第一章　符号与数学符号

人类是“从事构造化活动的动物”。人类总想给客观事物赋予某种意义和价值，利用符号认识新事物，使客观世界秩序化，创造出色的科学文化世界。

所有的人类活动都基于象征表示，而像数学这种具有抽象本质的活动就绝对需要符号体系。

§1　符　　号

符号，处处可见。学生的校徽，军人的肩章，商店的招牌，公路旁的警告标志，通讯中的旗语，密码，数学中数字、字母符号，物理中单位符号，化学中的元素符号，扑克牌上的“红桃”、“梅花”，计算机、雷达荧屏上的信号…正如著名语言学家皮埃尔·吉罗所说：“我们是生活在符号之间”。

一、什么叫符号

一般说来，“符号”就是某种事物的代号，它的意义是，采用一一对应的方式，把一个复杂的事物用简便的形式表现出来。当某事物作为另一事物的替代而代表另一事物时，它的功能被称为“符号功能”，承担这种功能的事物被称为“符号”。有时，当事者根据自己的主动性判断来断定某事物代表某事物。在这一瞬间前者也变成了“符号”。例如，人们看到“闪

电”就预感到“打雷”，“闪电”就成了符号；按照这种方式，人类实际上可以把所有的事物都冠以“符号”。

“所有事物都是符号”这种说法未免太广泛了。现代符号学中所谈的“符号”总是为传播一定的信息而用，不用于传播目的的自然标志则算不上符号。尽管人们经常说云是雨的标志，烟是火的标志，但是，符号学不承认它们的符号地位，因为这些“云”和“烟”并无意传播给我们一种信息，它和在地面上留下了形迹的作案犯一样。

不过，自然标志可以当作符号来使用，例如，电视机播放的气象图中的云，公安部门对于指纹的语言描述。

为了区别于自然标志，人们把符号定义为传播意识的一种意愿标志。一种符号就是一种刺激，也就是一种可感的实体，这种刺激为了传播而和另一种刺激的影像连在一起。因此，语言及语言的各种辅助手段，礼仪，交通讯号等等，都可看作是符号。

有些数学家，还对“符号”和“记号”两词加以区别。所谓记号是指一些我们感觉的外界事物，它们可以彼此区别，并且可以根根一定规则而对应于别的事物。例如，用耳感觉到的声音，用目感觉到的图形、文字，用收报机感觉到的电波等。

各记号之间有一种关系叫做同型关系或等价关系。这种同型关系满足三个性质：

自反性：每个记号与自己同型；

对称性：若甲与乙同型，则乙与甲同型；

可传性：若甲与乙同型，乙与丙同型，则甲与丙同型。

把同型的记号的共同性质抽象出来就可得到符号概念。

例如，伸出四个手指，绳上打四个结，墙壁上画四条横线

和碗里放四粒豆等，都可作为记号，它们可以看作是同型的，对这类记号进行抽象就可得到一个符号“4”的概念。

按照这种观点，记号被看作是具体的事物，而符号则是经过抽象而得的抽象的东西。在很多书刊中对符号与记号两词不加区分，有时，确实难以严格区分。

人类为什么要创造符号呢？事情很简单，当我们讨论或提到某一事物时，在一般情况下，不能使用该事物的本身，而必须使用表示该事物的符号。例如，当我们讨论到火山、地震或洪水时，我们不可能把“火山”、“地震”或“洪水”带到会议桌上；当我们说“没有见到过外星人”时，我们不可能拿出外星人。人们不得不处处创造符号、使用符号。

语言就是浸透到人类生活的各个角落的符号，是谁都知道且必须知道的符号。它是人们每天接触的最熟悉的符号体系；是人类接触到的各种各样符号体系中最典型的符号体系。人们常说“语言是传达思想的手段”，或者因为考虑到语言的感情功能而说“语言是表现、传达思想和感情的手段”。其实，语言不仅是“手段”，它还超越了手段的意义。“语言”同文化和思考问题的方法都有着深刻的联系。因此，有些学者提出，语言是“文化的象征，从而它能够规定思考方式”，语言是“文化的模式”，除掉具有作为手段和实用功能外还具有美学功能。

二、符号与人类的关系

符号源出于人类的“给予意义”的行为，即给予某种事物以某种意义，从某种事物中领会出某种意义。凡是人类所承认的“有意义”的事物均成为符号。

人类的“给予意义”的行为，在日常生活中，首先是依靠“语言”的使用进行的，通过“创造语言”给予某种事物以某种意义。“创造语言”似乎是非常崇高的事情。其实，这种“创造语言”的事情是所有的人都在不停的进行着的。最切近、简单的例子是日常生活中的“命名”行为。例如，某人为自己喂养的小猫起名为“小花”。为什么要起名字呢？当然是为了区别于其它的猫。这只猫对自己来说，具有不同于其它猫的特殊价值。由于赋予特别的名称，使其不能为其它所代替，从而完成了给予意义的活动。诗人、文学家、政治家等“为了打破日常语言的约定俗成的习惯”都不断地“创造语言”，并把这种由新鲜的语汇创造出来的意思，当作超越日常生活的手段。只要新的语汇具有某种意思（或能当作意义解释），那么它就是“符号”，已不再是按照习惯去表现旧有内容的“标志”，而是新诞生的符号。例如，“棋坛新星”、“东方巨龙”、“芦荡火种”等等。新的“符号”所产生的新内容给我们这个世界增添了新知识。这的确是一种创造活动，是“创造语言”的活动。

人们对于某些对象命名时，无须清楚了解它们的本来面目和性质。比如，“上帝”一词早已被创造出来，“上帝”的本来面目究竟是什么，也许始终弄不清楚。尽管如此，人们总想通过命名，假想某物的存在，并根据与自己的关系，给它规定一个位置。“上帝”这个符号是人们尝试抓住某个未知物，根据它与自己的关系给予它意义，把它引进我们这个世界中的产物。在自然科学中也常常引进某些未知的基本粒子或行星的符号，进行理性论证。近几年，人们把天空中某种“飞行物”命名为“不明飞行物”，关于“不明飞行物”的研究越来越受

到人们的关注。从宗教、艺术到科学的尖端领域，人们总是要给周围事物命名，赋予意义，给以符号，以便捕捉未知，研究客观世界，掌握客观世界。

现代，符号的概念已不再限于人类言语活动的一些标志，它已扩展到人类社会的很多方面。人文科学中的神话、宗教、文学等等，都被视为符号系统；自然科学的各个分支也几乎都各有其符号系统。不过，关于语言符号的研究，至今还是最重要的，因为，它"简直就是其他符号研究的基础和模式"。

人类使用以语言为中心的各种各样的符号，"肯定和维持已经诞生的人类文化秩序使之功能化"；"有效地处理新的事物，把握其意义和价值，并把它纳入秩序化的世界中"，以致创造新的科学世界。正如有些学者所说，人类是"使用符号的动物"。

三、符号形式与符号内容

任何符号都包括两个方面，即符号形式（能指）与符号内容（所指）。符号是传播意识的一种意愿标志，其核心就是用"某事物代表某事物"。任何符号总依赖于两个"某事物"之间相互依存的关系。人们就把这两项依次称作"符号形式（能指）"和"符号内容（所指）"。瑞士著名语言学家、现代语言奠基人索绪尔称前者为"声音形象"，而后者是"声音形象所表达的概念"，还有些学者把这两项称为"表达表面"和"内容平面"。例如，1949年南京"总统府"门楼上升起五星红旗，五星红旗是"符号形式"或叫"能指"，"南京解放"就是"符号内容"或叫"所指"。

符号的功能是用符号形式代表符号内容，其基础是"符号

形式”和“符号内容”之间的相互依存关系。缺少这两项里的任何一项，“符号”以至“符号功能”都不能成立。例如，鹦鹉模仿人类语言说“早上好”的时候，从表面看来使用了人类一致的符号，但是它使用这种符号并非出于同样的符号内容，所以不能说鹦鹉与人类使用的是同样的“符号”或“符号形式”，其实，它们都是一些无所指的或者说没有符号内容的事物，不应作为“符号形式”来处理。

另外，即使有内容，不给予表现形式，那它也谈不上是“符号”或“符号内容”。而且，就是在由于人的主动性解释，使“事物”“符号”化的构造过程中，也只有当某种意义（“符号内容”、“所指”）被解释后，这事物才化为“符号形式”。

“符号形式”与“符号内容”在逻辑上为相互依存的关系是十分明确的。但是，二者的相互依存关系在心理上并不一定总是对称的。在日常使用的语言中，“符号”一词的运用常常是指“符号形式”。这是由于在符号的两个侧面中，“符号形式”对我们来说，是可以以某种形式感觉到的对象，而“符号内容”则未必。由于符号形式属于符号“显眼的”侧面，所以它很容易与“符号”本身的存在联结起来，暗示“不显眼的”符号内容的存在，从而代表着“符号”整体。例如，“停止”这个符号形式，就属于这种情况。如果球赛正在进行，这“显眼的”“符号形式”就会暗示双方以“符号内容”——“停止球赛”。反之，用“符号内容”代表“符号”整体就困难了。

“符号形式”优先，最典型的例子是宗教性的象征符号。这种象征符号中的“符号内容”是在日常应用的符号的层次上难

以理解的。它超越了日常范围的体验。信徒们相信在作为"符号形式"的象征符号背后，存在着某种"符号内容"，并有意识地体验和把握它。纳粹分子的记号"卍"，其功能也近似于宗教性的象征符号。只要作为象征符号的"符号形式"被明确地规定下来，"符号内容"的存在就会充分得到暗示。

四、符号学

符号学，顾名思义，是有关符号的科学。

有关符号的思想，可以追溯到古代。古希腊最伟大的哲学家亚里士多德在《解释篇》中论及语言时，已涉及符号问题。他写道："由嗓子发出的声音是心灵状态的象征，写出的词句是由嗓子发出的词句的象征。同样，写出的文字，在所有的人那里不会一样，说出的话也不会都一样，尽管心灵状态在所有的人那里是一样的，以这些心灵状态为其意象的事物也是一样的。"这里，他已经谈到了声音、心灵状态和事物之间的关系；象征与符号是同义词。

稍后，斯多葛派哲学家们为了建立严密的三段论逻辑而明确讨论了符号问题。他们认为，符号有其发音的部分，有其被揭示的事物(它取决于人们的思维)，有其被指示的外部对象。它们是连在一起的，其中，声音和对象是有形的，而事物是无形的。这与人们现在所讲的能指、所指和指代对象三个概念已十分接近。

到了17世纪，伟大的哲学家、数学家莱布尼茨，全面地发展了斯多葛派哲学家关于符号的思想，他认为关于符号的科

学，应能排列符号，使其表达所思。他终生的事业之一是寻找一种通用语言，潜心研究《万能算法》。他的目标是建立一种符号和术语的体系，以便整理和简化逻揖推理的基本要素。他在数学中达到了自己的目标。在一切科学和数学中，他所建立的无穷小运算是符号和术语体系的极好范例。莱布尼茨被誉为历史上最伟大的符号学者之一。

19世纪，布尔的符号逻揖发展了莱布尼茨的《万能算法》。布尔在其《逻辑的数学分析》中，把数学方法引入了逻揖学，同时，提出不借助哲学和心理学就可以说明能指过程的一些数学程序。正是在这一道路上，美国哲学家、逻揖学家皮尔斯提出有必要建立一种新科学，这种科学论述符号的意指作用（能指与所指之间的关系），论述意指作用诸系统间的可调换性以及它们在物质范畴内的关系，这就是符号学。在他看来，符号学只不过是更广泛意义上的逻辑学的代名词，他写道：

“逻辑学，我认为我曾指出过，就其一般意义而论，只不过是符号学的另一种说法而已，符号学是关于符号的几乎是必然的和形式的学说。在把这门学科描述成‘几乎是必然的’或形式的学科的时候，我注意到，我们是竭力来观察这些符号的特征的，而且，根据这些观察，并借助于抽象活动的一种过程，我们已经到了可以对由科学才智使用的各类符号的特征进行十分必要的判断的时候了。”

他的《存在的曲线图》一书，使人看出，他曾试图建立一种广泛的符号学。

几乎在同一个时期，索绪尔把符号学设想为“研究社会生活中符号生命的科学”。他把言语活动看成是一个系统，而这个系统又不同于纯粹的逻辑系统。他写道：

“语言是一种表达观念的符号系统，因此，可以比之于文学、聋哑人的字母、象征仪式、礼节、形式、军用信号等等。它只是这些系统中最重要的。因此，我们可以设想有一种研究社会生活中符号生命的科学；它将构成普通心理学的一部分；我们称它为符号学。它会告诉我们符号是由什么构成的，受什么规律支配。因为这门学科还不存在，我们说不出它将会是什么样子，但是，它有存在的权利，它的地位是预先确定了的。语言学不过是这门一般科学的一部分，将来符号学发现的规律也可应用于语言学，所以后者将属于全部人文现象中一个非常确定的领域。”

这里索绪尔明确了语言学和符号学之间的关系，并预示了语言学研究将会对普通符号学的发展做出贡献。

人们一般认为，皮尔斯与索绪尔是现代符号学的奠基人。索绪尔着重符号的社会功能，皮尔斯着重符号的逻辑功能。这两方面是密切相关的。

有关符号的一般理论是在本世纪初才出现的。这种理论最初以普通语义学的名称，首先引起了逻辑学家们的兴趣。索绪尔的设想，很晚才得到实施，那是1964年，法国著名符号学家罗兰·巴特写出了《符号学要素》一文。

1969年，国际符号学研究协会成立，这标志着符号学研究进入了一个新阶段。之后，欧、美、前苏联及日本等国均有不少人开始研究符号学；很多学科都采用了符号学的研究方法；符号的分类研究，导致了符号学研究的各个分支，使人们看到符号学在不少领域具有广阔的前景。

目前，数学符号学，还是亟待开垦的“处女地”。

§2 数学符号

数学的世界是一个符号化世界。

使用符号是数学史上的一件大事。符号和公式等人工语言的制定是最伟大的科学成就，它在很大程度上决定了数学的进一步发展。数学语言“就像一座灯塔，照亮了黑暗中的巨大地域，照亮了自然的未被揭索的秘密。”

一、数学语言

数学的语言是由一些符号和记号组成的语言。

数学概念本身是抽象的，无影无踪的，为了把数学概念传播出去，就必须借助于一种具体的，使人能感受到的可代用物。这种不得不被无休止地使用的代用物就是数学符号。

“符号是交流与传播数学思想的媒介”。世界各国都有各自的语言，汉语、英语、法语、德语、俄语……，但数学语言可以世界通用，它是“国际性的，是唯一完全国际化的语言。”全世界的人，凡是受过初等教育的人，都认识下列符号语言，不需要再翻译：

$$(a+b)(a-b)=a^2-b^2$$

$$\cos 2\alpha=2\sin\alpha\cos\alpha$$

$$\triangle ABC\cong\triangle A'B'C'$$

………

形式主义学派曾提出一种看法：数学对象就是符号本身，数学的命题则是由一定法则组成的符号系列，这些符号和符号系统可能有直观的涵义，但这些涵义并不属于数学。现代形式

主义学派代表柯恩还提出，数学应当被看作一种纯粹的纸上符号游戏，对这种游戏的唯一要求就是它不会导致矛盾。

尽管这些看法具有片面性，但它从一个侧面反映了数学语言的符号化特征；他们的研究工作曾产生过积极的作用，推动了数学的发展。

数学的对象决不是符号本身。一百多年前恩格斯曾指出："纯数学的对象是现实世界的空间形式与数量关系。"根据百多年的发展变化，现在徐利治先生进一步指出，数学是"实在世界的最一般的量与空间形式的科学，同时又作为实在世界中最具有特殊性、实践性及多样性的量与空间形式的科学。"由于"数学思想超越时代与地域，在社会的、政治的以及经济的条件之外进行，……作为社会和经济条件的产物的自然语言不适合于这种目的，因而不得不发明了一种为进行抽象思维的符号语言。"尽管这种语言是抽象的，但数学为生活的各种哲学提供了新的基础。

数学语言是按下列不同方向改革自然语言的结果：(1) 简化自然语言；(2) 克服自然语言中含糊不清的毛病；(3) 扩充自然语言的表达范围。

自然语言有不方便之处。例如，把两数和的立方公式用自然语言来表达，该是：

"两数和的立方等于第一数的立方、第一数的平方与第二数的积的 3 倍、第二数的平方与第一数的积的 3 倍以及第二数的立方这四项的和。"

这样用文字叙述，很繁琐，实在令人讨厌。如改成数学语言：

"$(a+b)^3=a^3+3a^2b+3ab^2+b^3$"

就大大缩短了语言表达的"长度"，并且还呈现出对象本身具

有的对称性，数学语言必须“从冗长的自然语言中解放出来。”

自然语言有含糊不清的地方，比如，“机关”可以表示政府部门的各种职能机构；可以表示阴谋诡计：“机关算尽太聪明，反误了卿卿性命。”还可表示各种密室、金库等大门上的防盗设施：“罪犯中了机关”。数学需要准确而清楚的语言，其中每一个符号，每一个由符号组成的式子只有一个意思。

数学语言和自然语言之间的本质区别之一是变元的使用。由于使用了各种变元，数学语言能够很好地表示一般规律，极大地扩充了语言表达的范围。

可见，把数学语言叫做符号语言是有道理的。尽管自然语言也使用一定的符号，但它与数学语言是有本质区别的。“数学语言对任何人来说，不仅是最简单明了的语言，而且也是最严格的语言”。当代，“数学不仅是事实和方法的总和，而且是用来描述各门科学和实际活动领域的事实和方法的语言”。

二、名字

一种语言如果满足两个条件，它就足以准确描述某类对象：(1)这类中的每个对象都有相应的语言表达——名字；(2)不同的对象都有不同的名字。

如果不满足第一个条件，那么语言就太贫乏，不足以描述这类对象；如果不满足第二个条件，那么语言就是含糊不清的。

可见名字在语言中的重要地位，它是与讲话中所表示的对象联系在一起的。一般地，人们把语法中称为专有名词的那些词（如“上海”、“函数”、“太湖”）叫做名词。对象是其

名字的代表者或意义，它既可以是具体的实际对象，也可以是数学抽象物（如方程 $x^2-6=0$，$\triangle ABC$，指数函数的导数），还可以是任意的概念、性质、关系、过程、事件或事件系列等。

同一个对象可以有许多不同的名字，这些名字按不同的方法描述给定的对象。例如：

记号 $5-3$，$8\times 2-14$，$(7+13):10$ 都可作为数2的不同的名字；

记号 $\{1, 2, 3, 4\}$，$\{3, 2, 1, 4\}$，$\{1, 2, 1, 3, 1, 4\}$ 都是同一个集合的名字；

等式 $5-3=(7+13):10$ 表明，该式左、右两边的名字表示同一个对象。

但是，在数学中用作名字的每一个符号，表示的对象不能多于一个。

必须区分对象及其名字。在一个句子中遇到一些对象的名字，这一句子讲述的是对象，而不是它的名字。例如，“将数66分成两半”这句话，是要把记号66所表示的对象——数66分成两半，结果应该等于33。如果把这句话误解为“将记号66分成两半”，那么结果将是数字符号 6 。

我们知道，人们对于某些对象命名时，无须清楚了解它们的本来面目和性质。在数学中情况也一样。一般说来，有可能理解了名字的涵义，但是对其代表者却什么也不知道，只知道它是由其涵义确定的。例如

“方程 $x^5-5x^3+1=0$ 的最小实根”

这一名字涵义是很清楚的，尽管我们不知道解五次方程的公式。利用这个名字的涵义，可以计算这个根到任意的精确度。

“有可能理解了名字的涵义，而并不知道其代表者是什么，只知道它是由其涵义确定的。”这种可能性是人类有能力利用语言去认识新事物的基础。

由此可见，应该将“没有涵义”与“没有意义”这两个术语区分开来。例如，在自然数范围内，名字“方程$x+4=3$的根”是没有意义的，或者说这样的对象是不存在的，但这时这个名字却有着清楚的涵义：它是这样的一个数，将它代替方程中的x以后等号左边与右边得到同一个数的名字。同样地，在实数范围内，名字$\sqrt{-4}$没有意义，但它有涵义——平方等于-4的数。

另外，一个名字有没有意义，与所讨论对象的范围有关。当对象的范围扩充时，本来只有涵义的名字也可能获得意义。例如，名字$\sqrt{-4}$，在将实数集扩充到复数集时就可得到确定的值。

三、变元

在数学中，始终表示同一对象的记号叫做常元。因此，人们把各种数学对象的名字称为常元，因为每个名字始终表示同一个对象。例如，名字“6”，“sin”（表示正弦函数），“N”（表示自然数集），“$\sqrt{3}$”等等都是常元。

在数学中，常常会遇到这样的符号，它表示某一集合的“任意的对象”。

例　求证任一自然数与其后继数的积恒被2整除。

考虑到自然数集N的无限性，我们不可能用列举法对每个自然数逐个进行证明。这时可以先取“任一自然数n”，再对这个数“n”分两种情况讨论：

“当n为偶数时，$n(n+1)$是偶数；当n为奇数时，$(n+1)$是偶数，因此$n(n+1)$是偶数”。

在上述证明过程中用到了符号n，它表示自然数集中任意的元素。如果用自然数的具体元素的名字来代替，就可证得这个具体元素符合命题的结论（例如，若$n=3$，则$3(3+1)$是偶数）。

可见，上述推理过程中的符号n起着具体名字的代表者的作用，就像在自然语言中代名词起着名词代表者的作用一样。在数学中，与这种符号n相类似的符号就叫做变元。

这就是说，变元是一种符号，它可以表示某一集合的“任意的对象”。这个集合是由可以代替变元的元素所组成的，它叫做这个变元的值集或变域。通常要求它是非空的。

数学表达式的性质与变元的变域密切相关。例如，要把x^4-4分解因式，则需根据变元x所在的变域，分别对有理数域，实数域，复数域逐个讨论。因此，使用变元时，始终要注意变域。

变元是一个非常基本的概念，表面上看来又是极其简单的概念，以致有的数学家讲：“变元的概念并不是逻辑上必需的，可以根本不必引入。”事实上，变元是“数学语言的元素”，“不论在数学中、数理逻辑中乃至日常语言中都占着非常重要、非常关键的地位。”“要澄清数学中某些模糊不清或容易混淆的概念，就离不开变元。苏联数学教育家斯托尼亚尔曾指出：“传统教学法的根本缺点之一，是它没有形成学生关于数学语言的最重要因素——关于名字和变元的概念。”

目前，很多人已接受这样的观点，所谓变元，应该是指符号不应该指事物或指数量；严格地讲，变元是一种符号，它可以用以代替某类符号中的任意一个符号。通常，人们说“变元

可表示任意一个变域中的事物”，数学家罗素也说“变元所指是含混的，因而是不确定的事物”。这些说法都是不够确切的。所谓“任意”，所谓“不确定”，应该就符号而言，不应该就外界的事物而言。因此，很多读者把变元看作“变量”，以为是外界的量的一种，这种看法是不够妥当的。

关于“变量”一词，自从17世纪引入数学以来，它的意义一直是不清楚的。一提到变，自然要牵涉到时间，而时间在数学中还没有很好地定义过。从而，“自变量”这个提法本身也是有缺点的，因为它必定依赖于时间而变，不可能脱离时间而“自变”。有些学者主张废弃变量这个词，而将自变量改称为“Argument”（德文），但是，它至今没有恰当的译名，不应再译成自变量是肯定的。还有人主张干脆将“自变量”、“因变量”分别改称为“first entry”，“Second entry”，而这两个词也没有很好的译名。因此，人们还只能按习惯不断地用“变量”，“自变量”，“因变量”等词。

变元可代替其变域中的任何符号，反过来，它的位置便可由变域中的任何符号来填入。因此，变元基本上就是空位。在国外的文献中常把变元称为“位置保持者”（“Placeholder”（英文）、“Platzhalter”（德文）），它反映了变元的实质，非常富于表现力。

我国第一届华罗庚金杯少年数学邀请赛中，有一题是：将0，1，2，3，4，5，6这七个数字填在圆圈和方格内，每个数字恰好出现一次，组成只有一位数和两位数的整数算式。问填在方格内的数是几？

$$\bigcirc \times \bigcirc = \square = \bigcirc \div \bigcirc$$

现在我国小学数学教学中经常使用这类带"小空盒子"或者"小窗户"的语言。这样做有好处，借助这种"语言"甚至能进一步得到代数中一系列的公式，例如：

$$\square + \bigcirc = \bigcirc + \square$$

$$(\square + \bigcirc)^2 = \square^2 + 2\square\bigcirc + \bigcirc^2$$

美国小学数学教育现代化活动的领导人戴维斯曾断言，从"小窗户"过渡到字母很自然而且也比较容易。

这样做，可以得到关于变元的正确观念。其根本原因就在于"变元基本上就是空位"。

但是，变元和空位毕竟不相同。数学中不可能长期使用带"小窗户"的语言。对这种带"小窗户"的式子要进行数学推演是困难的、不方便的。因此从长远说来，必须引进变元的概念及符号，否则数学是不能向前发展的。

在某些符号组合中出现的古怪变元是不能用其值替代的。例如，符号组合

$$\sum_{n=1}^{n=4} \frac{1}{n(n+1)}$$

看来好像含有变元n，但如果用 3 来代替n，就得到一个无意义的符号组合

$$\sum_{3=1}^{3=4} \frac{1}{3(3+1)}$$

事实上，"$\sum_{n=1}^{n=4} \frac{1}{n(n+1)}$"是一个完整的符号，它表示和式

$$\frac{1}{1\cdot(1+1)}+\frac{1}{2\cdot(2+1)}+\frac{1}{3(3+1)}+\frac{1}{4\cdot(4+1)}+$$
$$+\frac{1}{5(5+1)}+\frac{1}{6(6+1)}$$
其中并不含有变元。

严格地讲，这种古怪的变元根本不是变元，但通常则说，它以约束形式引入符号组合，称为约束变元，例如：

记号“$\sum\limits_{n=1}^{n=6}$”将变元n“约束”在符号组合“$\frac{1}{n(n+1)}$”中（在$\frac{1}{n(n+1)}$中，n是自由变元，它可以取任一自然数值。）

在公式$\lim\limits_{x\to 2}x^2=4$和$\int_1^2 x^2dx=\frac{7}{3}$中，$x$也是作为约束变元引入的，在第一个公式里它约束于算符“$\lim\limits_{x\to 2}$”，在第二个公式里它约束于“$\int^2\cdots dx$”。

在公式$x^2=4$和$x<10$中，x是自由变元。

在某些符号组合中，同一个字母既可作为自由变元，又可作为约束变元，例如：$\int_1^x x^3dx$。

如果某变元是约束变元，那么它可以换成其他的任一符号，而不改变表达式的涵义。例如记号

$$\sum_{n=1}^{n=6}\frac{1}{n(n+1)}\text{与}\sum_{K=1}^{K=6}\frac{1}{k(k+1)}$$

有同一涵义——它们表示同一个和式。如果在表达式中变元是自由变元，那么用另一个字母替代它时，表达式就改变了涵义。例如，表达式“$3x+6$”与“$3y+6$”是不同的。

在数集中取值的变元称为数值变元。在现代数学中，不仅要应用数值变元，还广泛应用点、向量、函数、集合以及命题的

变元。例如，我们考察命题：

“如果 x ，那么数$3a$被 9 整除” 。

如果这里的 x 用命题“数a被 3 整除”代替，就得到一个真命题：

“如果数a被 3 整除，那么数$3a$被 9 整除” 。这里，变元x的值是一个命题，这样的变元 x 就是一个命题变元。

各种类型的变元的应用极大地扩展了数学语言的表达力。

当然“变”与“不变”总是相对的。例如，在

“函数$f:x\longrightarrow\{x\}$除了在整数点上间断外，在其余一切的点 x 上连续”

这个命题中，字母f用作常元，即作为一个确定的函数的名字。而在命题

“闭区间上的连续函数f，在该区间上总能取得最大值与最小值”

中，字母f用作变元（它是一个全称量词，表示闭区间上的任意一个连续函数）。

在中学数学里，经常用形如“函数$f(x)=x^2$”的式子来表示“函数$f:x\longrightarrow x^2$”的意义。在这些式子里，字母f也是用作常元。

四、数学语言表达的分类

数学符号从其作用上可分为以下四类：

1. 元素符号。表示数或几何图形的符号称为元素符号。例如：

数字符号 0，1，2，3，4，5，6，7，8，9；

表示数的字母 a，b，c，…，x，y，z，…；

某些特定的常数 e，π，i；

三角形的边 a，b，c；

三角形的角　A，B，C；

另外，还有△，∠，▱，⊙等。

数字符号及e、i、π等为常元，其它字母符号则往往用作变元。

2．关系符号。表示数、式、形之间的关系的符号称为关系符号。例如：

$=$，$>$，$<$，$\backsim$，$\cong$。

3．运算符号。表示按照某种规定进行运算的符号称为运算符号。例如：

$+$，$-$，$\times$，$\div$，$\cdot$，a^n，$\sqrt{\ }$，Σ；

sin，cos，tg，ctg，arcsin，arccos，arctg，arcctg，log，ln，lg；

lim，f，f'，y'，dx，$\int$；

行列式符号及矩阵符号。

4．辅助符号。为了便于表达和运算，数学中还引进了一些符号用于表示某些特定的式子或某种特定的意义，例如：

Δ　一元二次方程的判别式；

$n!$，Max，min；

括号　(　　)，[　　]，{　　}。

表示三角形全等的符号（s,s,s），（s,a,s），（a,s,a）。

当然，以上所列举的，仅是一些常见的基本符号。

数学符号也可按数学科学的领域来划分，在各个领域中都要运用一系列专门的符号。从中可以排出基本(原始)符号的某一集合，这些符号类似于外语字母表的符号，人们也称为字母，它们组成为字母表。例如，代数语言中的字母表为

$\{0, 1, \cdots, 9, a, b, \cdots, x, y, z, +, \cdot, =, <, (,)\}$

代数中使用的其他符号都由这个字母表中的符号确定。

字母表中的符号是原始“材料”，按一定法则用它来构造各种语言表达式——类似于自然语言中的词和句。

数学语言和自然语言一样，每个表达式都有两方面内容：语义内容和语法内容。所谓语义内容，是指符号表达式与它表示的对象之间的关系，它指示了符号表达式的内在涵义。例如，$a+b=b+a$这一表达式的语义内容是：对于“+”这种运算来说，元素的先后次序不同并不影响运算结果。所谓语法内容是指符号表达的形式结构。

数学语言的一切表达式（广义的，也包括传统的“代数式”和“公式”）是字母表中的符号的序列，但是，不是所有的符号序列都是数学语言的表达式。作为数学语言的表达式必须是按一定法则构造出来的、有意义的符号序列，比如，“2+1”是表达式，而符号序列“2 + +”不是表达式。因为，符号“+”表示施行在两个数上的运算，所以在语言表达式中，它只能占两个数或两个变元的符号之间的位置。

所有语言表达式的集合可以分成最简的和复杂的两类。最简的由一个字母，一个数字（或者数字的有限序列，如6636）组成；复杂的，除了字母和数字之外还包含其它符号，

对于不含变元的表达式，可以分成两类：

(1) 不包含关系符号的表示数的表达式。

例如，“2 + 1”与“3”，它们都表示数。

(2) 包含关系符号“=”，“<”或“>”之一的、表示命题的表达式。

例如，符号序列“2 + 1 = 3”也是语言表达式，但是它的意思和“2 + 1”及“3”的意思完全不同，这是一个表示命题的表达式。

对于包含变元的表达式也适于做类似的分类。为了这个目的，我们观察、比较下列表达式：

(1) $7+x$；

(2) $7+x=10$，$7+x<10$，$7+x>10$。

从语法观点看，(2)的每个表达式和表达式(1)的区别在于它包含某个关系符号（“=”，“<”或“>”）。显然，这样只从语法上理解它们的区别是不够的，还必须弄清这些表达式的意思，即采取语义处理。我们利用函数的思想来分析。比如，假定变元x的取值范围是集合$A=\{1, 2, 3, 4, 5\}$，表达式(1)是表示数值变数的数值函数，或者集合A到集合$B=\{8, 9, 10, 11, 12\}$上的映射$A\to B$的数值形式；(2)的每个表达式是命题形式，它表示数值变数的逻辑函数（谓词），或者集合A到集合$\{T, F\}$〔由“T”（真）和“F”（假）两个元素组成的集合〕上的映射$A\to\{T, F\}$，这可由定义这个函数的表直观地表示出来：

x	$7+x$	$7+x=10$	$7+x<10$	$7+x>10$
1	8	F	T	F
2	9	F	T	F
3	10	T	F	F
4	11	F	F	T
5	12	F	F	T

函数的思想把语言形式的不同范畴结合起来了。数值形式和命题形式都表示函数：第一个是数值函数，第二个是逻辑函

数。

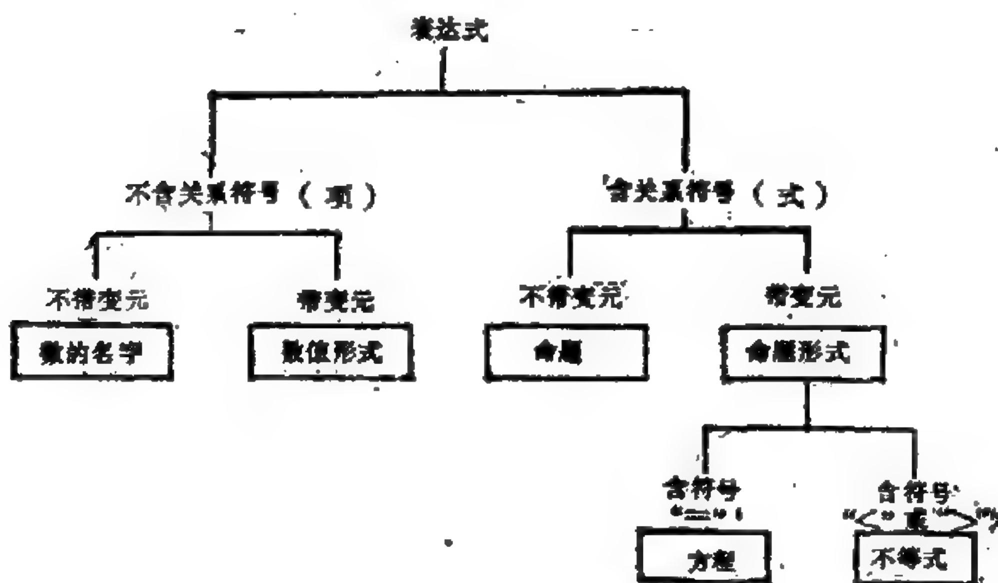
第一种情况（“$7+x$”），函数的定义域和值域都是数集，其余情况（“$7+x=10$”，“$7+x<10$”，“$7+x>10$”），定义域是数集，而值域是集合$\{T, F\}$。

为此，我们可把含变元的表达式划分成两类：

(*a*) 不含关系符号的，为数值形式的，而且确定一个数值变元（或多个数值变元）的数值函数的表达式；

(*b*) 包含关系符号（等于、小于或大于）的，为命题形式的（给变元代入值后，就变成命题），而且确定一个数值变元（或多个数值变元）的逻辑函数的表达式。

由上面的分析可以引出如下数学语言表达式的分类表：



这里所用的表达式这个术语，在传统的学校术语中叫做“代数表达式”。这个分类可以提供一个基础，以便在学校代数教学中使用名词术语时更有秩序，而且更好地理解表达式的语

义。

很多其它数学领域的语言表达的分类，均可参考上述分类法。

存在一些特殊的表达式，它虽然含有变元，但也表示命题而不是命题形式。例如，公式

$$\lim_{x\to 2} x^2 = 4 \text{ 与} \int_1^2 x^2\, dx = \frac{7}{3}$$

它们含有变元x，但表示真命题。

这两个表达式中变元x是作为约束变元出现的，不是自由变元。

§3　数学符号发展简史

整个的数学符号体系只有四百来年的历史，实际上，大多数数学符号还远远没有四百年历史。数学符号的发展大致可分为五个阶段。

一、建立自然数和分数的符号体系，特别是引入位值制计数法及零的特殊记号

现在世界通用的记数法是十进位值制记数法；运用的数字符号是印度·阿拉伯数码1，2，3，…，9，0。所谓位值制，就是一个数码表示什么数，要看它所在的位置而定。如23和32使用两个同样的数码，但23中的2放在十位上，表示20；32中的2放在个位上就表示2。十进制显然和人类的十个指头有关。恩格斯曾在《反杜林论》中讲过："数和形的概念不是从其他任何地方，而是从现实世界中得来的。人们曾用来学习计数，从而用来作第一次算术运算的十个指头，可以是任何别

的东西，但是总不是悟性的自由创造物"。

古代，数的概念虽然早已发生，但表示数目的符号的发展是相当迟缓的，它经历了一个漫长的演变历程。"如果把世界上所有不同的数码写法罗列出来，那将要写成一本几百页的大书。"现在国际上运用的印度·阿拉伯数码是印度人发明的，它本身的演变也有一段漫长复杂的历史。最初印度人用梵文(印度古代文字)的字头表示数码，且各地的写法也不完全相同。经过几百年的演变，在8世纪时传入阿拉伯。当时印刷术还未发明，书籍全用手抄，字体因人因地而异，出入很大。因此，在不同地区就出现了不同的书写形式，这就是所谓的东阿拉伯数字和西阿拉伯数字。12世纪时，两种形式又逐渐合流，开始传入欧洲。欧洲人只知道这些数码是从阿拉伯国家传来的，所以又称为阿拉伯数码。14世纪，中国的印刷术已经传到欧洲。在1480年英国有些印刷本书籍中数码已相当接近现代的写法。到1522年，英国阿斯托的书中所用数码才和现在的写法基本上一致。以后渐渐稳定下来。

位值制记数法是千百年人类智慧的结晶，它可以同字母的发明媲美，两者都是用少数简单的记号来代替复杂难记的符号。

古埃及人很早就用10进记数法，但却不知道位值制，每一个较高的单位是用特殊的符号来表示的。例如111，埃及象形文字记作



其中，第一个符号表示1，第二个符号表示10，第三个符号表

示100。

中美洲的马雅人懂得位值制的道理，但用的是20进制。巴比伦人也知道位值制，而用的是60进制。使用位值制而又是十进的，以中国人为最早。

在14世纪以前，中国的计算工具一直是一种细长的小棍棒，这小棍棒是用竹子、木头或骨制成的，这就是算筹。

算筹是什么时候开始使用的，年代已遥不可考。但有资料证实，在公元前2、3世纪时，算筹的运用已达到相当纯熟的地步。

用算筹表示数目，有纵横两种方式：

纵式	𝍠	𝍡	𝍢	𝍣	𝍤	𝍥	𝍦	𝍧	𝍨
横式	𝍩	𝍪	𝍫	𝍬	𝍭	𝍮	𝍯	𝍰	𝍱
	1	2	3	4	5	6	7	8	9

记数时，个位常用纵式，其余纵横相间。例如6728表示作𝍮𝍦𝍪𝍧。6708表示作𝍮𝍦 𝍧，空一格的地方表示零。个位是零也能表示出来，如6720表示作𝍮𝍦𝍪。由于各位数字纵横相间，所以不致看错。位值制的关键是零号，没有表示零的方法，位值制就不完备。巴比伦人位值制的思想起源很早，但缺乏适当的零号。

完整的位值制记数法，必须有表示零的记号。零号的创设是位值制的必然物。尽管空位的方法出现如此之早，但代之以用圆圈表示的零号“〇”却迟迟难以产生。迄今为止所发现的第一批载有零号的文字，同时出现在公元683年柬埔寨和苏门答

腊的碑文上。那里用ⓒ·ε表示605，用ⓒ○∀表示608。前者用“·”表示零，后者与近代的零号一致。前一种表示法在印度使用得比较早。至于用“○”表示零，考虑到东南亚各国文化曾受到中、印两国的重大影响，因此一些科学史家倾向于认为它是公元4世纪左右产生于中、印两国的边界一带。

“零”这个字，原来并不表示空无所有的○。根据《说文解字》中的解释，“零”是作为“零头”，如“零丁”、“零星”、“零碎”都是这个意思。105读作“一百零五”，原来是指一百之外还有一个零头五。后来“○”也就读作零了。符号“○”则演变为扁圆的“0”。

世界上也有不少民族懂得零的道理，然而系统地研究、处理和介绍零，还是以印度人的功劳最大。

分数概念的产生，最初并不是由除法来的。分数被看作是整体或一个单位的一部分。后来，在运算过程中也产生了分数，它表示两个整数的商。

三千多年前埃及纸草书就已经有分数，但缺乏合适的分数符号，计算非常复杂。巴比伦人用60进分数，运算也麻烦。中国《九章算术》是世界上系统地叙述分数的最早著作，比欧洲大约早1400年。

中国古代用算筹做除法，所得的商列在上面，被除数（称为“实”）在中间，除数（称为“法”）在下面。这种排法已接近现在的记法，分子在上，分母在下，只是带分数的整数部分排在上面。

1881年在印度西北边界的巴哈沙利附近，挖得写有文字的桦树皮。大概是8、9世纪时转抄3、4世纪时算术书的残页，

称为巴哈沙利残简。该残简中，将$\frac{1}{3}$写成$\begin{matrix}1\\3\end{matrix}$（缺分数线）而$1\frac{1}{3}$写成$\begin{matrix}1\\1\\3\end{matrix}$，也是把带分数的整数部分写在上面。人们根据足够的理由推测，印度的分数记数法是中国传去的。12世纪印度最突出的数学家婆什迦罗的《丽罗娃提》中也采用这种写法。如

$$3+\frac{1}{5}+\frac{1}{3}\text{写作}\begin{array}{|ccc|}3&1&1\\1&5&3\end{array}\text{通分后变成}\begin{array}{|ccc|}45&3&5\\15&15&15\end{array}$$

后来传到中亚细亚，也将分子写在上，分母写在下。目前所发现的最早的分数线是在12世纪时阿尔·哈萨的著作中出现的。按照他的写法，

$$\frac{3\ 3\ 2}{5\ 8\ 9}\text{表示}\frac{2+\dfrac{3+\dfrac{3}{5}}{8}}{9}$$

阿拉伯文的书写是从右到左的。斐波那契《算盘书》沿用他们的习惯，式子也是从右到左，整数部分写在分数的右边，将$12\frac{1}{2}x$写成“radices$\frac{1}{2}$12”。这是在欧洲最早出现的分数线。

分数线没有马上被大家采用。14世纪中叶还有用$3\,\overline{\rule{0pt}{0pt}}5$表示$\frac{3}{5}$的。为了节省地方，英国棣么甘推荐用$a/b$表示$\frac{a}{b}$。这种记法在18世纪末叶已经出现。

二、建立代数的符号体系

代数学发展的关键是要建立一套有效的符号体系。

代数符号是在悠久的岁月中经过不断改良、选择和淘汰的结果。内塞尔曼曾于1842年指出代数符号体系演变的三个阶段

——文字代数、简写代数与符号代数。

最初是文字代数。解题不用简写记法，也不用任何符号，而是用地道的散文形式写成。3600年以前的古埃及的纸草书上，用象形文字表示一次方程，在公元300年以前整个代数都是文字代数。到了希腊时代，代数学获得重大的发展。代表人物是丢番图（公元246—330年），被誉为代数学的鼻祖。他的重要贡献之一就是对希腊代数引进简写记法，他用简写文字表示三次多项式。他还用字母来表示未知元和一些运算，这是近世符号代数的嚆矢。不过，在世界的很多地方文字代数还相当普遍地持续了好几百年。公元 9 世纪，阿拉伯人阿尔·花拉子模的《代数学》中，一切算法都用文字语言来表达。例如把“$x^2+10x=39$”说成“一个平方数及其根的十倍等于39”。尤其是西欧，15世纪以前，基本上都是文字代数，直到后来，才开始有一些零星的简写记法。

15世纪，阿拉伯人已开始在方程中使用印度·阿拉伯数码，但形状和现在还有很大的区别，而且按阿拉伯人的习惯，式子是从右向左读的。

符号代数是16世纪才开始在西欧出现的。但发展缓慢，直到17世纪中叶才广泛流行。在符号体系上使代数产生最大变革的是法国数学家韦达（1540—1603年）。他第一个比较有意识地、系统地在代数中引入符号体系。他的名著《分析方法入门》一书对符号代数的发展有不少贡献，被认为是一部最早的符号代数著作。其中，他用辅音字母表示已知元，用元音字母表示未知元。

对多项式的系数加以修饰，用字母表示一般的系数。还使用过现在的“+”号和“－”号。在韦达之前，人们一般用不

同的字母表示一个量的各种幂，韦达用同一个字母，并适当地加以说明。他明确指出代数和算术的区别，他说算术是与数打交道的，而代数是研究表达式和解方程的。但是他缺乏等号与乘号。他的符号体系还是混乱的，言词与缩写混在一起。例如，他写的

A cubus+B planum in A, aequatur

D Solido

就是现在的

$A^3+3BA=D$ 或 $x^3+3Bx=D$

英国数学家雷科德于1557年出版了一本代数学，书中首次使用了现代的等号。过去也曾有人使用平行线段来作等号，但是写得太长。

1637年，笛卡尔的《几何学》问世，他提出用字母表中后面的字母x，y，z表示未知数，用前头的字母a，b，c表示已知数，他所给出的代数方程的形式和现在的写法已基本一致，只是缺少幂的符号，他把x^2、x^3写成$x\cdot x$、$x\cdot x\cdot x$。现代的代数符号系统主要是采取笛卡尔的符号。

三、与微积分学的产生相联系的符号的发展

微积分学不是凭空产生的，它经过了相当长时期的酝酿，最后在17世纪末叶，经牛顿、莱布尼茨的手而完成。正如恩格斯所说，微积分是“由牛顿和莱布尼茨大体上完成的，但不是由他们发明的”。微积分的计算方法在牛顿、莱布尼茨之前已散见于各国的著作中，其中积分的思想，早在希腊时代已经萌芽。牛顿和莱布尼茨最大的功绩是将两个貌似不相关的问题联系起来，一个是切线问题（微分学的中心问题），一个是求积

问题（积分学的中心问题），建立了两者之间的桥梁——现今所谓的“牛顿、莱布尼茨公式”。牛顿、莱布尼茨各自引进了一些导数、微分、积分等有关符号，但牛顿引进的符号，基本上早已淘汰。而莱布尼茨建立的符号一直使用到现在，没有什么实质性的变化。

四、集合论与数理逻辑符号在数学中的发展和渗透

在现代数学中，“集合论已构成全部数学的基础”。它的概念和方法几乎渗透到数学的各个分支乃至其它很多自然科学部门。数理逻辑，又称为符号逻辑，是研究推理，特别是研究数学中的推理的科学。

现代，集合论和数理逻辑的符号正在逐步地向数学领域发展和渗透。例如：

集合符号：A、B、C；

$\in$——属于；$\cup$——并；$\cap$——交；

T——真命题（真）；F——假命题（假）；

$\varnothing$——空集；$\subseteq$——包含；

$\overline{A}$——集合A的补；

$A\times B$——集合A乘以集合B的直积；

$A^2=A\times A$——A中元素的一切可能序偶的集合；

$\underset{x}{M}[p(x)]$或$\{x|p(x)\}$——具有性质P的一切x的集合；

$\overline{x}$——命题x的否定；

$\wedge$——合取；$\vee$——析取；

$\Longrightarrow$——推出；$\Longleftrightarrow$——等价；

$\forall$——全称量词；$\exists$——存在量词。

这些记号大体上是19世纪中叶以来，由意大利数学家皮亚诺引进的。

这些符号促使数学语言更加简便，更便于推理。例如，关于函数一致连续的定义：

“函数f称为在区间〔a，b〕上一致连续，如果对于任意的$\varepsilon>0$，总存在这样的$\delta>0$，使对于区间〔a，b〕上任意的x_1，x_2，$|x_2-x_1|<\delta$，必有不等式$|f(x_2)-f(x_1)|<\varepsilon$成立”。

现在可以写成：

“函数f称为在区间〔a，b〕上一致连续，如果：

$\forall\varepsilon>0\quad\exists\delta>0\quad\forall x_1\in[a,b]\quad\forall x_2\in[a,b]$

$|x_2-x_1|<\delta\Longrightarrow|f(x_2)-f(x_1)|<\varepsilon$”

这样的写法容易构造所给性质的否定，只要通过改变量词与否定不等式即可：

“函数f称为在区间〔a，b〕上不一致连续，如果：

$\exists\varepsilon>0\quad\forall\delta>0\quad\exists x_1\in[a,b]\quad\exists x_2\in[a,b]$

$|x_2-x_1|<\delta\wedge|f(x_2)-f(x_1)|\geqslant\varepsilon$”

五、高速电子计算机的出现刺激了数学符号的新的发展

计算机语言不同于一般的数学语言，其符号必须便于输入计算机。特别地，由于程序的需要，所用的符号不能带下标和指数，不能应用通常的幂和根的符号。例如，在计算机程序以及专门为程序所写的数学文本中，x^2写成$x\uparrow2$，$\sqrt[3]{x}$写成$x\uparrow$（1.0/3.0）。

随着数学的发展，数学符号也将不断地发展。

第二章　数学符号的动力作用

数学在发展，数学符号也随之创新、发展；另一方面，数学符号的创新又推动了数学的发展。

数学发展的动力究竟在哪里？毫无疑问，它来自社会生产实践及科学技术发展的客观需要，但是除了这种外部推动力之外还有内在的动力因素。正如希尔伯特所说：

“在每个数学分支中那些最初、最老的问题肯定是起源于经验，是由外部现象世界所提出。……但是，随着一门数学分支的进一步发展，人类的智力，受着成功的鼓舞，开始意识到自己的独立性。它自身独立地发展着，通常并不受来自外部的明显影响，而只是借助于逻辑组合、一般化、特殊化，巧妙地对概念进行分析和综合，提出新的富有成果的问题……”

数学符号的使用是推动数学发展的内在动力因素之一。“数学的一切进步都是对引入符号的反应。”在数学里，有人把17世纪叫做“天才的时期”，也有人把18世纪叫做“发明的时期”，这两个世纪数学为什么会有较大的发展呢？原因之一，就是这两个世纪大量创用了数学符号。

历史表明，数学符号与数学方法有密切的关系。数学上对一般方法论的关心出现于16至17世纪之间，它正是由于代数符号体系的建立而引起的。目前，关于数学方法论的研究正方兴未艾，研究数学符号对数学发展的动力作用是研究数学中的发现、发明与创新等法则的重要课题之一。

诚然，数学符号的动力作用可简单概括为一句话："没有数学符号，就没有现今的数学"。不过，这样讲未免太笼统了，作为数学方法论的研究课题之一，我们必须结合数学的发展过程作一些具体的剖析。

§1　符号对数学发展的影响

数学符号演化的自身规律表明，数学的符号化必须适应数学体系发展的需要，"一种合适的符号要比一种不良的更能反映真理"。符号的优劣直接影响数学发展的速变。在数学的发展过程中，一方面对符号的改革与日俱增，另一方面符号的改进又加速了数学学科的发展。

欧洲在阿拉伯数码输入之前，使用罗马数码。在这种计数法中用

Ⅰ，Ⅱ，Ⅲ，Ⅳ，Ⅴ，Ⅵ，Ⅶ，Ⅷ，Ⅸ，Ⅹ，L，C，D，M分别表示

1，2，3，4，5，6，7，8，9，10，50，100，500，1000

它不是进位制的，一个简单的数要写成长长的一串，这种笨拙的记数法在12世纪以前盛行于欧洲，有的国家直到16世纪还在使用。那时，"做加减法已相当困难，会乘除法就可以称为专家了。"数学史中记载了当时算术教科书中的一个乘法实例——计算235×4的过程：

235写成CCⅩⅩⅩⅤ。乘上4（Ⅳ），第一步是将CC，ⅩⅩⅩ，Ⅴ分别重复写4遍：

CC　CC　CC　CC

XXX　XXX　XXX　XXX

V　V　V　V

第一行共有 8 个C，将 5 个C缩写成D（500），第二行10个X缩写成C（100），第三行缩写成XX（20），于是简写成

D　C　C　C

C　X　X

X　X

再进一步合并，得到结果$DCCCCXL$（XL是40）。

这只是用一位数去乘的情形，如果是多位数乘多位数，其复杂的程度不难想见。加法并不比乘法简单多少，乘法只是重复写若干遍，而加法要逐个数有多少个I，多少个V，多少个 X，……，然后再缩写成所求答案。至于分数运算，德文里有这样的谚语，形容一个人已经陷入绝境，束手待毙，就说他已"掉到分数里去"。

这种情形严重地阻碍了数学的发展。正因为如此，用印度·阿拉伯数码代替罗马数码就势在必行了，而印度·阿拉伯数码的使用就明显地促使数学得到迅速发展。

阿拉伯数码传入我国，最早是13、14世纪，但迟迟不被采用。19世纪末期，有些书本中，正式采用了阿拉伯数码，但不是横写而是直写，如

2 3 5 8　写成　2
3
5
8

直到本世纪初，才慢慢改用现在的写法。

数学史家分析，我国迟迟不用阿拉伯数码，主要原因可能是我们长期使用算筹式数码记数法，也是十进位值制，和阿拉

伯数码的记数法效果相同，而汉字一、二、三、四、……笔画也很简单易写，一时看不出阿拉伯数码的显著优点。欧洲中世纪用的是笨拙的罗马数码，和阿拉伯数码记数法相比，显得十分落后，因此容易接纳新的记数法。

无疑，迟迟不用阿拉伯数码，对我国的数学的发展是有一定影响的。

微积分的出现是整个人类历史的一大喜事。恩格斯高度重视微积分在自然科学中的作用，他指出："只有微积分才能使自然科学有可能用数学来不仅表明状态，并且也表明过程：运动。"科学的微积分符号的作用对发展微积分成果、推动后世自然科学的发展极为重要。正如拉普拉斯所讲："数学分析的语言，是所有的数学语言中最完善的语言，而且语言本身就成为新发展的有力工具。特别是那些被构思出来的种种必要概念，往往是许多新算法的起点。"如果没有完整的微积分符号，自然科学就不可能发展到如今的水平。

众所周知，微积分"大体上是由牛顿和莱布尼茨完成的"，现在的微积分符号基本上是沿用莱布尼茨所创设的。

牛顿也创设了一套符号。

1669年，牛顿完成了第一篇微积分论文，题为《运用无穷多项方程的分析学》。其中，牛顿提出一种积分法，把变量的无限小增量叫做"瞬"，并记为"0"。

例如，设有一条曲线$y=f(x)$，其下的面积为$Z=ax^m$（m为有理分数），当横坐标x获得一个无限小增量，即瞬0时，有新坐标$x+0$，并产生面积增量，即面积瞬oy，新面积为：

$$Z+oy=a(x+o)^m$$

由二项定理

$$Z+oy=a(x^m+mx^{m-1}\cdot o+\frac{m(m-1)}{1\cdot 2}x^{m-2}\cdot o^2+\cdots)$$

考虑到$Z=ax^m$，并在等式两边除以0，得

$$y=amx^{m-1}+a\cdot\frac{m(m-1)}{1\cdot 2}x^{m-2}\cdot 0+\cdots$$

略去含有 0 的项，得到

$$y=max^{m-1}$$

这就是相应于面积Z的纵坐标y的表达式。这个结果表明，若面积$Z=ax^m$给出，那么构成这个面积的曲线为$y=amx^{m-1}$。反之，如果曲线是$y=max^{m-1}$，那么，它下面的面积就是$Z=ax^m$。

在这里，牛顿不仅给出了求一个变量对另一个变量的瞬时变化率的普遍方法，而且通过证明面积可以由求变化率的逆过程得到，揭示了微积分的基本性质。但是，其中有不少含混的地方。例如“0”是不是零？牛顿认为不是，既然这样，为什么在运算中可以略去含有 0 的幂的项呢？牛顿没有给出合乎逻辑的说明。

1671年，牛顿整理了他六年来关于微积分的研究工作，写成《级数法和流数法》，这是牛顿在数学方面的代表作。此书在1736年才正式出版。其中，牛顿引进了他独特的符号和概念：他把随时间变化的量，即以时间为独立变数的函数称为流量，用字母表的后几个字母v、x、y、z来表示；而把流量的变化速度，即变化率称为流数；或简称为“速度”、“迅度”，用加小点来表示。如$\dot{x}$表示x的流动率，$\dot{y}$表示y的流动率，余类推。因为时间是所有流动量的自变量。$\dot{x}$，$\dot{y}$，…相当于莱

布尼茨的$\frac{dx}{dt}$，$\frac{dy}{dt}$，…。他保留“瞬”的概念，并仍用“0”表示。

稍后，牛顿在另一著作《曲线求积论》中重新解释他的符号：在下列符号

$$\overset{||}{x},\quad \overset{|}{x},\quad x,\quad \dot{x},\quad \ddot{x},\quad \dddot{x}$$

中的每一个是前一个的流动率（导数），是后一个的流动量(原函数）。

其间，他还指出了确定x^n的流动率的方法：

当x“由流动”变成$x+0$，x^n变成$(x+0)^n$，展开成级数：

$$(x+0)^n=x^n+n0x^{n-1}+\frac{n^2-n}{2}0^2x^{n-2}+\cdots$$

两边减去x^n，就得到函数的增量

$$(x-0)^n-x^n=n\,0\,x^{n-1}+\frac{n^2-n}{2}0^2x^{n-2}+\cdots$$

为了避免“略去含有0的项”，牛顿接着考虑自变量增量(0）与函数增量的比。它等于

$$1:nx^{n-1}+n\frac{(n-1)}{2}0x^{n-2}+\cdots$$

当增量消失的时候，它们的“最后比”就是

$$1:nx^{n-1}\text{ 或 }\frac{1}{nx^{n-1}}$$

这个“最后比”不是别的，它是x的流数与x^n的流数之比。这种方法，除了极限的说法外，大致和现在教科书的方法相同。

综上所述可见，牛顿在创建微积分时也不得不考虑符号问题。但是，牛顿更多的是关心创立微积分的体系和基本方法，而莱布尼茨似乎更关心运算公式的建立与推广，力求建立微积

分的规范，即法则和公式的系统。莱布尼茨对符号的关心超过牛顿，他是历史上最大的符号学者之一。他所创设的微积分符号远远优于牛顿符号，这对微积分有极大影响，正像印度·阿拉伯数码的采用促进算术发展一样。

有一段数学史对于阐明我们这节的主题是有益的。

1684年莱布尼茨发表微分法，1686年发表积分法。以发明的时间来说，牛顿是先行者（早10年），但他的流数术到1687年才以几何形式发表出来，而《流数术》本身直到他死后9年（1736年）才印刷出来。莱布尼茨比牛顿较早公布（早3年）。为此，英国数学家和欧洲大陆数学家之间曾发生一场不幸的论战。其实，这种争论没有实际价值，反倒使英国的微积分发展受到影响。

大陆派的学者在接受了莱布尼茨优越的符号以后，经过伯努利家族、欧拉、达朗贝尔、拉格朗日、拉普拉斯等人的进一步工作，很快地获得丰硕的成果，渗透到各个数学部门中去。

英国的情况如何呢？苏格兰的克累格（J.Craig）在1685年采用了莱布尼茨的概念和符号。三十多年后，由于英国人狭隘的民族偏见加上对牛顿的盲目崇拜，放弃这种符号而改用牛顿的“流数术”，迟迟不肯接受大陆的成就，因此，其进展相应地落后了。

数学符号的优劣对数学发展的影响，从这个历史教训中，便可略见一斑。

§2 速记符号引申出新分支

为了有利于传播数学思想，有利于运算和推理，在数学中

必须构造很多速记的符号。例如幂的符号“a^n”、阶乘的符号“$n!$”、行列式以及矩阵符号等等都是速记的符号。诸如此类的速记符号，不仅帮助人们理解已有的理论，简化运算、推理步骤，还能帮助人们看清更深刻的数学关系，建立更为重要的理论，乃至引申出新的数学分支。正如希尔伯特所说：“数学中每一步真正的进展都与更有力的工具和更简便的方法的发现密切联系着，这些工具和方法同时会有助于理解已有的理论，并把陈旧繁杂的东西抛到一边。数学科学发展的这种特点是根深蒂固的，因此，对于个别的数学工作者来说，只要掌握了这些有力的工具和简单的方法，他就有可能在数学的各个分支中比其它科学更容易地找到前进的道路。”

美国数学家M·克莱因曾说过：“行列式和矩阵完全是语言上的改革，对于已经以较扩展的形式存在的概念，它们是速记的表达式。”这里，有必要追本溯源，看看它们的形成简史。

行列式概念起源于对线性方程组的研究。我国古代数学家解线性方程组时，用算筹把未知数的系数排成方阵，在对二元和三元线性方程组用消元法求解的过程中，导致了一个二阶方阵或三阶方阵与一个数的对应，这就蕴含了行列式的思想。

最早引入行列式概念的，是17世纪日本数学家关孝和。他在1683年发表的《解伏题之法》（意思是“解行列式问题的方法”）一书中，已经谈到行列式和它的展开问题。

在欧洲，最早研究线性方程组并论及行列式的是莱布尼茨。

首先用行列式的方法解线性方程组的是英国数学家马克劳林。他的记法不太好。稍后，瑞士数学家克莱姆在他的《线性

代数分析导言》一书中给出了用行列式解线性方程组的法则，就是现今的“克莱姆法则”。他还提出了行列式展开法则。1764年，法国数学家裴蜀把确定行列式每一项的符号的手续系统化，并证明：系数行列式等于零是齐次线性方程组有非零解的条件。

后来，荷兰数学家范德蒙把行列式理论与线性方程组求解问题相分离，对行列式理论作了系统的逻辑阐述。

现在行列式理论已经广泛应用到很多数学分支。

矩阵与行列式一样，也是由研究线性方程组引行起的。

我国《九章算术》中，把线性方程组

$$\begin{cases} 3x+2y+z=39 \\ 2x+3y+z=34 \\ x+2y+3z=26 \end{cases}$$

表示成矩形的方阵：

未知数不用符号表示，只将各个系数用算筹依次罗列出来，文字采用直排，而且阅读时是从右到左的。如果按现代的习惯来表示，这个算筹方阵就是现代教科书里的增广矩阵。《九章算术》中用加减消元法解线性方程组实际上就是对矩阵

元素按一定的规则进行加减和乘以某个数的运算。可惜，在我国未能发展成完整的矩阵理论。

从逻辑上说，矩阵的概念先于行列式的概念，而在历史上次序正好相反，矩阵概念是从行列式的研究中引申出来的。行列式的研究开始于18世纪中叶以前。行列式包括一个数字方阵，通常总是涉及这个方阵的值，就是由行列式的定义所给出的值。然而人们从行列式的大量研究工作中发现，对于很多问题，方阵本身都可以研究和使用，不管行列式的值是否与该问题有关。方阵本身就称为矩阵。矩阵这个词是英国数学家西勒威斯特首先使用的，他希望引用数字的矩形阵列而不能再用行列式这个词，从而提出矩阵这个词来。不过，他仅仅涉及到一些方形的矩阵。

英国数学家凯利首先发表一系列文章论述矩阵的基本概念、性质及运算等。因此，他被誉为矩阵论的创造者。他说："矩阵…或是直接从行列式的概念而来，或是作为一个表达方程组：

$$\begin{cases} x' = ax + by \\ y' = cx + dy \end{cases}$$

的方便的方法而来的。"就这样他引进了矩阵：

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

它代表了这个方程组的主要信息资料。

数学家C.V.牛松(C.V.Newsom)曾讲过：

"和许多经典科学相比，现代的社会科学与自然科学发现了数学的符号式阐述具有重要意义，这种意义表现在它具有把

从社会和自然的宇宙中得到的数据联系起来的能力，以及因而还具有对带有科学重要性的本质问题做出回答的能力。"

行列式和矩阵极其明显地具备上述能力。

什么是行列式理论？塞尔维斯脱（sylvester）说："它是代数上的代数，这是一种使我们能够把代数运算组合起来并预言结果的演算，这种情况，就像代数本身能使我们不必进行具体的算术运算也行之有效的情况是一样的。所有的分析最终都必须以这种形式作为自己的外衣。"

下面再谈矩阵：

在解析几何里，二次曲线的方程

$$ax^2+2bxy+cy^2+2dx+2ey+f=0$$

与矩阵

$$\begin{pmatrix} a & b & d \\ b & c & e \\ d & e & f \end{pmatrix}$$

密切相关，它可以认为是上述二次曲线的矩阵，而该矩阵的行列式

$$\begin{vmatrix} a & b & d \\ b & c & e \\ d & e & f \end{vmatrix}$$

就是这二次曲线的一个不变量。

这里，矩阵把曲线方程的各项系数"联系起来"，用以表示二次曲线，而且呈现出明显的效果。

在工程技术、经济工作中，有些问题的解决需要考察多方面的因素，从整体上反映其数量关系，如电子网络系统，经济

规划，商品产销关系等等都可用矩阵把有关的信息指标“联系起来”。比如有S个产煤地区：A_1，A_2，，…，A_s需要销售到n个地区：B_1，B_2，…，B_n，则一个调运方案就可用一个矩阵

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{s1} & a_{s2} & \cdots & a_{sn} \end{pmatrix}$$

来表示。其中a_{ij}表示产地A_i运到销地B_j的数量。

在图论中，把各种具体事物的线路图抽象为图，图又可以借助于矩阵把有关指标联系起来，从而得到图的矩阵表示。

在对策论中，可以把竞争双方的各种决策方案及其相应的竞争效果用一个矩阵“联系起来”，便于谋求最优策略。

如此等等，不胜枚举。

这种能力，反映了数学符号的生命力。正如美国数学史家D.J.斯特洛伊所讲，合适的数学符号“带着自己的生命出现，并且它又创造出新生命来”。数学发展史证实了这种观点。当代行列式理论已经广泛应用到很多数学分支，矩阵已成为很多现代数学分支不可缺少的工具。矩阵理论本身也有了极大发展，本来，矩阵的元素都是实数，后来发展为可以是复数，并且可以是许多别的量。19世纪末和20世纪初的研究已经专门针对元素属于抽象域的矩阵性质。另外，由于积分方程理论的推动，人们还研究了无限阶矩阵。

“只要世界上还存在着数学家，革新就不会停止”。新的速记符号的创设还会引申出新的分支。

§3 符号化导致新学科诞生

我国古代的数学成就辉煌，明朝以前在许多方面一直遥遥领先于西方各国。我国是最早使用十进制、位值制记数法和零的文明古国之一。最早使用分数和小数，最早系统地总结了几何学的基础知识，最早发现勾股定理，最早掌握线性方程组的一般解法和高次方程根的近似解法，最早研究了不定方程、一次同余式方程组，并得出了著名的"中国剩余定理"，等等。但是，自明朝中期以后，我国的数学水平却一直停滞不前。到清朝后期，我国的数学水平就大大落后于西欧各国了。原因是多方面的，其中令人十分痛心的一个重要原因就是：我国古代数学忽视了数学符号的使用。

近代的数学家们都十分重视数学的符号化，符号化使数学本身以及其它以数学为主要工具的科学的面貌发生了革命性的巨大变化。

代数学是怎样成为独立学科的？

近代，代数学是以研究各种代数结构的性质为中心。然而，在19世纪以前，代数一直被理解为关于方程的科学，它的发展是和方程分不开的。

代数和算术的主要区别，就在于前者引入未知量，根据问题的条件列出方程，然后解方程求出未知量的值。1873年，华蘅芳和英国人傅兰雅合译英国华里司《代数术》。一开始就指出："代数之法，无论何数，皆可以任何记号代之…"说明所谓代数，就是用符号来代表数字的一种方法。

代数学发展的首要一步就是引进符号代表数字，用符号代

替文字叙述。在第一章§3，我们已经看到在这方面功劳最大的该推韦达，他第一个比较有意识地、系统地使用符号表示数字和算式，并对字母符号进行运算，使“代数”开始正式成为一门学科而独立出来。正如著名德国数学家克莱因所说：“代数学上的进步是引进了较好的符号体系，这对它本身和分析的发展比16世纪技术上的进步远为重要。事实上，采取了这一步，才使代数有可能成为一门科学。”

古希腊的数学在几何学方面曾取得辉煌成就，但在代数学方面的发展却较缓慢，这在很大程度上是由于缺乏适当的符号体系。在古埃及、古巴比伦、古印度和古代中国数学中符号化程度都很不彻底，因而未能发展成具有近代形态的代数理论。

符号化不仅导致代数学的独立，还直接刺激解析几何的诞生。

大家熟知解析几何是借助坐标系用代数方法研究几何问题的一门学科。

17世纪，法国数学家笛卡尔的著作《几何学》是解析几何学的起点。笛卡尔是一位杰出的近代哲学家和近代生物学奠基人，又是出色的物理学家和数学家。他努力寻求发现真理的方法，把方法论作为他一切工作的首要对象。他于1628年写成第一部著作《思想的指导法则》，1634年写成第二部著作《世界体系》。1637年，他出版了《更好地指导推理和寻求科学真理的方法论》，包括三个著名附录，上述《几何学》就是其中之一。

当时“几何学”一词，并不专指现在的“几何”而言，它和“数学”是同义语。正像我国古代“算术”和“数学”是同义语一样。

其实，坐标法的思想可以追溯到远古的希腊。阿波罗尼斯（前260—前170年）研究圆锥曲线的时候，曾引用了两条正交直线，作为一种坐标。他取圆锥体底面直径作为横坐标，过顶点的垂线作为纵坐标。稍后，天文学家依巴谷采用经纬度来表示星球的位置，其实质就是用两个坐标来确定一点的位置。到14世纪，法国数学家奥力森陈述一种坐标几何，用坐标来确定点的位置，这是从天文、地理坐标到近代解析几何学的过渡。他的著作对笛卡尔的工作有一定的影响。但是，刺激笛卡尔创立解析几何的直接因素还是韦达的符号代数。笛卡尔强烈地意识到代数的巨大威力，他把代数看做是进行推理，尤其是对未知量进行推理的有效方法。在笛卡尔看来，欧几里得过分强调证明的技巧，过分依赖图形，对每个问题的证明都需要经过一个新的特殊的方法才能解决。这不仅是“笨拙和不必要的”而且使得几何“失去科学的形象”，而代数又太受法则和公式的束缚，影响人们思维的灵活性。然而，笛卡尔觉察到符号代数具有一种能使思维“机械化”的可能，这种思维机械化会使运算步骤变得简单，达到以最少思维获得最佳结果的效能。正是由于笛卡尔看到了代数方法的巨大力量。出于一种对方法论的强烈兴趣，笛卡尔才着手把代数用于几何的伟大工作，写出上述《几何学》。

变量的思想首先就是笛卡尔在《几何学》中引入的，但他没有使用变量的这一术语。他称一些量为“未知和未定的量”，就相当于现在的变量。

笛卡尔《几何学》第二卷中有这样一个例子：

如图2-1，设直尺GL的一端固定在G点上，可以绕G点旋

转，$AK \perp GA$，有一个三角板CKB的边BK贴在AK直线上，上下移动。使直尺通过三角板BK边上的固定点L，求GL与三角板CK边（或延长线）交点C的轨迹。

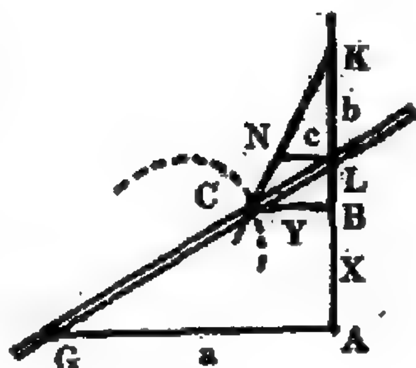


图 2-1

笛卡尔选择直线 AB 作为量点的位置的标准，以 A 作为始点（用现代的术语来说就是以AB作坐标横轴，A 点作坐标原点。）

作$NL \parallel CB \parallel GA$。他写道："因为$CB$和$BA$是两个未知和未定的量，分别命它们为$y$和$x$"。

又命$GA=a$，$KL=b$，$NL=c$。因$c : b=y : BK$，故

$$BK=\frac{b}{c}y,$$

$$BL=\frac{b}{c}y-b,$$

$$AL=x+\frac{b}{c}y-b.$$

$$\begin{aligned}\text{又}\quad CB : BL &= y : \left(\frac{b}{c}y-b\right)\\ &=GA : AL\\ &=a : \left(x+\frac{b}{c}y-b\right)\end{aligned}$$

故$\frac{ab}{c}y-ab=xy+\frac{b}{c}y^2-by$，所求的轨迹方程是

$$y^2=cy-\frac{cx}{b}y+ay-ac$$

笛卡尔说这是一条双曲线。

这里我们可以看到笛卡尔是怎样引入变量的。

无疑，变量的引入对数学的发展有很大影响，为此，恩格斯对笛卡尔的工作给予极高评价，他说："数学中的转折点是笛卡尔的变数，有了变数，运动进入了数学，有了变数，辩证法进入了数学，…"

数理逻辑是怎样产生的？

数理逻辑是数学和传统逻辑相结合的产物，在其中，人们使用数学符号来研究数学各领域公共使用的逻辑推理，简单地讲，就是把传统逻辑加以符号化，所以又名符号逻辑。

所谓传统逻辑主要是指亚里士多德逻辑，尤其是经过中世纪的演变一直沿用到19世纪乃至现代的那种逻辑。亚里士多德早在《前分析篇》中研究了直言三段论这种推理的形式和规律，例如

推理的内容	推理的形式
所有的人都有死	所有M是P
苏格拉底是人	所有S是M
∴苏格拉底有死	∴所有S是P

由于这种逻辑是以推理的形式，因而也是以命题的形式为其研究对象，所以人们又称之为形式逻辑。

数理逻辑的创始人是莱布尼茨。他的目标是建立一种符号和术语的体系，以便整理和简化逻辑推理的基本要素。在他的《组合术》中提出了建立普遍的推理演算系统的计划和设想。他试图把代数学的符号和演算应用到推理过程，这就开创了把数学和形式逻辑结合起来的先例，因此，一般认为他是数理逻辑的创始人。

到了19世纪中叶，数理逻辑沿着两个不同的方向发展。

第一个方向是逻辑代数化的方向。这个方向的工作，主要是把初等代数和形式逻辑相结合。这方面的突出人物是布尔（1815—1864年）。他在1847年发表了论文《逻辑的数学分析，论演译推理的演算》，1854年又发表了《思维法则的研究，逻辑和概率的数学理论的基础》。布尔正式提出改革传统逻辑的主张及具体方案。其基本思想与莱布尼茨的主张一脉相承，就是试图通过语言的符号来达到逻辑的严格化，因此，他致力于把推理过程的思维法则表述成符号语言的演算法则。因此，大家都承认他是继莱布尼茨之后的数理逻辑的第二个创始人。

布尔所从事的就是莱布尼茨想做而并未完成的工作，即仿照数学的方式来发展逻辑。他的成果便是今天有名的布尔代数。为了能在逻辑学中应用代数方法，他强调外延的逻辑，即类的逻辑，这样，他就能把逻辑的语言翻译成代数的语言了，其具体做法为：

逻辑的语言 —翻译成→	代数的语言
事物	个体变数X，Y，Z，…
概念	类变数X，Y，Z
论域	全类　1
自相矛盾的概念	空类　0
交叉概念	类的交　X·Y
并列概念	类的和X+Y

如此，逻辑推理法则就可以表述成代数演算法则了，例如：

逻辑规律	代数公式
矛盾律：任何一个事物不能同时是X，又是非X	$X\cdot(1-X)=0$
排中律：任何一个事物或者是X，或者是非X	$X+(1-X)=1$

在这种基础上，布尔证明了代数演算和逻辑演算之间有完全的形式上的类似。

布尔代数的建立，是数理逻辑宣告诞生的一个主要标志。布尔代数的两部分内容——集合代数与命题代数，都是现今数理逻辑的基本部分。布尔代数当初纯粹是一种理论研究成果，后来发展出一种开关代数，它在组合电路、电路网络等方面有极大的应用。

第二个方向是数学逻辑化的方向。这个方向的工作主要是和数学基础问题的研究相联系。由于数学的发展，特别是从解析几何与微积分开始的近代数学的发展，再加上两千多年来关于第五公设问题的研究，在18世纪末，数学的逻辑证明的严格性问题被提到十分突出的地位。当时，整个数学基础问题的研究，实质上是围绕着数学的原始概念的精确性和数学的逻辑证明的严格性进行的。

在这方面的突出人物之一是弗雷格，他是当时最大的数理逻辑家。他不仅完整地发展了数理逻辑的一个基础部分——命题演算，而且几乎很完备地发展了数理逻辑的另一个基础部分——谓词演算，特别是，他还把数理逻辑和数学基础问题紧密联系起来。其实，数理逻辑的整个基础到弗雷格手里已经接近于完成，因此，有人把他看作是数理逻辑的第三个创立者。可惜，弗雷格的符号和历来相传的、当时使用的、迄今使用的都完全不同，以致他的书当时根本未受到人们的注意，他的学说当时无人理睬。

在这方面的另一位突出人物是意大利数学家皮亚诺，他不仅从逻辑方面而且从语言方面对数学基础问题进行研究。他力图把数学建立成一个自足的形式系统。为此他对古希腊流传下

来的公理方法加以改进。为了保证理性思维发挥作用，防止不自觉地诉诸经验和直观，皮亚诺又提出了两个方法：(1)不用自然语言，采用一种符号语言。(2)把数学的推理写成符号公式。这实质上就是把数学语言和数学推理加以形式化。数学理论从此以一种新的形态出现：

第一，每一数学理论只包含少数几个原始概念，可以用原始符号来代表；

第二，每一数学理论只包含少数几条原始命题，可以用符号写出的公式来表述；

第三，数学理论的逻辑展开，完全变成符号公式的数学推演。

现今所沿用的数理逻辑记号大体上是由皮亚诺制定的，他的关于自然数论的三个原始概念和五个公理一直沿用到现在，成为自然数论的出发点，是数学理论形式化的一个杰出的成果。

皮亚诺可以作为数理逻辑的完成者，但他没有好好总结。严格说来，应该说完成者是罗素。这已是20世纪的事。罗素继承皮亚诺的研究，而且进一步系统地把符号逻辑与古典数学紧密结合起来。罗素和怀特海合著的《数学原理》，是直到当时为止数理逻辑的成果的总结，是数理逻辑发展史上一部划时代的著作。

之后，数理逻辑进入蓬勃发展时期，其种种成果已不能再用"数理逻辑"的总标题。

"几乎数学的每一分支都靠一种符号语言而生存"。几乎数学的每一分支都离不开符号化。

§4　符号美的魅力推动数学发展

“那里有数学，那里就有美。”数学有其自身的美。著名的法国数学家阿达玛写过一本《数学方面的发明心理学》，书中曾发挥了数学家、天体物理学家彭加勒关于数学发明创造的学说，他们一致认为，数学事实间的最佳组合往往是依靠“审美直觉”来作出的，没有美感的人，不可能成为数学发明家。著名英国数学家哈代曾讲过：“现在也许难以找到一个受过教育的人对数学美的魅力全然无动于衷。”

数学美的魅力是数学发展的深层驱动力。数学语言是由符号组成的，数学美的魅力主要来自数学的符号美。

一、数学符号美的主要标志

数学美的几个主要标志——简单性、统一性、对称性与奇异性等在数学符号方面均有反映。

“数学的语言不仅是最简单和最容易理解的语言，而且也是最精炼的语言。”简单性是数学符号美的最突出的标志。

数学符号的简单美决不是指简易、单薄、初等，而是要用简单的公式概括大量的事实，因而，这种简单同时就显得深远，显出数学符号之美。哲学家罗森在评价爱因斯坦时指出：“在构造一种理论时，他采取的方法与艺术家所用的方法具有某种共同性；他的目的在于求得简单性和美（而对他来说，美在本质上终究是简单性）。”例如，爱因斯坦的质能关系式：

$$E=mc^2$$

（其中E为能量、m为质量、c为真空中的光速）深刻地揭示了

微观、宏观、宇观无数质能变化现象的规律，但其式子非常之简单。优秀的诗词讲究用最少的字来表达最丰富的内容，而这公式用字之少，表达的内容之丰富，远非任何一首诗词所能比拟的，给人以深刻的美的享受。在我国出版的纪念爱因斯坦的邮票上，印上了这个简洁完美的公式，它代表了这位大科学家对人类贡献的精华。

莱布尼茨用“$\int f(x)dx$”这一简洁的符号表达了积分概念的丰富的思想，刻画出“人类精神的最高胜利”，因此，有些数学家把微积分比作“美女”。

事实上，自然界的许多现象可以归结为数学的一个公式、一个方程或一个函数关系式，都充分显示了数学符号的简单美。

与简单性有着密切关系的统一性，也是数学符号美主要标志。英国数学家、菲尔兹奖获得者阿蒂亚曾讲过：“数学中的统一性和简洁性的考虑，都是极为重要的。因为研究数学的目的之一，就是尽可能地用简洁而基本的词汇去解释世界。归根结底，数学研究是人类智力活动，而不是计算机的程序。如果我们希望能把人类所积累起来的知识一代一代地传下去，我们就必须努力地去把这些知识加以简化和统一。”

直线、圆、椭圆、双曲线与抛物线五种曲线在直角坐标系中，分别有各自的方程。然而，引进极坐标之后，这些曲线竟统一于一个简单的极坐标方程

$$\rho=\frac{ep}{1-e\cos\theta}$$

连中学生也为之赞美。

对称是一种美。从古希腊的时代起，对称性就被认为是数

学美的一个基本内容。著名德国数学家和物理学家韦尔说：“美和对称紧密联系”。韦尔与段学复都曾写过关于“对称”的专著，令人读之赏心悦目。许多数学符号都呈现出对称美。例如，对称多项式；对称矩阵；代数式化简时的共轭因子；代数方程虚根的成对出现；线性方程组的克莱姆法则等等都给人一种对称性的美感。数学上还有大量的公式，如海伦公式、平均不等式、赫尔德不等式、二项式定理、韦达定理、莱布尼茨公式，分部积分公式等等也都具有某种对称性。

奇异的东西能给人以美感，奇异之极是极美。英国哲学家培根说，没有一个极美的东西不是在和谐中有着某些奇异。数学中的奇异美也表现在数学符号方面。

众所周知，数学中有五个最突出的数，即1，0，i，π，e。这五个数是最具有代表性的。

“1”是一切实数的出发点，通过它及自然数对可构造全部实数；

“0”是所有实数中唯一的中性数；

“i”是虚数的基本单位；

“π”和“e”是两个地位特别的超越数。在§5我们将看到，π的历史尤为悠久，要弄清它的底细是不容易的。e作为近代发现的超越数，成为普遍使用的自然对数的底。

“1”，“0”代表算术，i代表代数，π代表几何，e代表分析，这么五个具有显赫地位的数却被欧拉用一个极简单的式子

$$e^{\pi i}+1=0$$

把它们联系起来了。这个公式揭示了数学各分支之间的内在联系和统一性，真是神奇绝妙。在数学家看来，这的确是美的象

征之一。

研究素数分布所服从的规律，一直是数学家们感兴趣的事。为此，人们致力于寻求素数在自然数中平均分布的信息。

对自然数n，设A_n表示整数1，2，3，…，n中素数的个数。高斯在实验、观察的基础上提出猜想：比值A_n/n“渐近等于”$1/\ln n$。就是说：

$$\lim_{n\to\infty}\frac{A_n/n}{1/\ln n}=1,$$

记为：$\frac{A_n}{n}\sim\frac{1}{\ln n}$。

素数分布的平均状态竟能用对数函数来描述，这是一个很引人注目的发现。因为，两个似乎完全无关的数学概念在事实上竟有如此紧密的联系，这是很令人奇怪的。果真，一百多年之后，人们证实了高斯的猜想。它成为整个数学中最著名的发现之一。

在一段时期内，人们普遍认为，函数的连续性保证了它的可微性。到了18世纪后期，数学家们认为，连续函数至少在某些点处可以微分，然而，德国数学家维尔斯特拉斯却在1872年的一次讲演中给出了一个处处连续而又处处不可微的函数：

$$f(x)=\sum_{n=0}^{\infty}b^n\cos(a^n\pi x)$$

式中x为实数，a是奇整数，$0<b<1$，$ab>1+\frac{3\pi}{2}$。

当时，曾有人批评这样的函数是“无意义的”，是“离奇古怪的函数”，是“病态函数”，是函数中“不健康的部分”，破坏了“数学天堂的优美”。于是有人“痛惜”，有人“惊恐”，也有人“厌恶”。出现这种情况是不足为奇的，因为数学美是“理智美”、“内在美”、“逻辑美”，要领悟数学美需要有一

定的数学素养，还要有个认识过程。历史证明，正是由于这种“奇异”函数的提出，人们才认识到连续函数不一定有导数和不连续函数也可求积分等重要结论，促进了数学分析的发展。

二、美的追求导致发现

科学美给人以美感；科学家的美感直接影响科学创造。

科学创造，除了反映世界，还来源于对美的追求。“美的概念在核对结果和发现新规律中被证明是非常宝贵的”，衡量一个科学理论是否成功，不仅有实践（实验）标准，传统逻辑标准，而且也有审美标准。对科学美的追求，常常为科学家们提供重要的线索和有力的手段。科学史家库恩指出，在新理论的建立中美的直觉的重要性有时可以是决定性的。麦克斯韦在建立电磁理论的过程中，依据法拉第的实验研究结果，把电磁理论方程写成

$$rotE = -\frac{1}{c} \cdot \frac{\partial H}{\partial t}$$

和

$$rotH = 0$$

当然，两个等式的右边是不对称的。麦克斯韦从数学美考虑出发，把第二个方程改成

$$rotH = \frac{1}{c} \cdot \frac{\partial E}{\partial t}$$

他这样做，当时并没有任何实验依据，但却使电磁理论在数学的符号形式上更优美，并可导致一系列令人陶醉的结果。后来，改写的方程被实验证实了，电磁理论决定性的一步就是这样跨出的。

韦尔曾经说：“我的工作总是力图把真和美统一起来，但

当我必须在两者中挑选一个时，我总是选择美。对于韦尔的中微子两分量相对论波动方程，物理学家们三十多年不理睬它，但韦尔凭美的直觉相信它。结果，韦尔对了。

对数学美的追求是狄拉克一生科学活动的内在驱动力之一。他认为："如果物理学方程在数学上不美，那就标志着一种不足，意味着理论有缺陷，需要改进，有时候，数学美要比与实验相符更重要。" 因为数学美反映了基本的自然规律，而实验的是否符合可能与一些细节有关。1927年，狄拉克认为克莱因方程还很不完美，无论从逻辑上还是从数学上都有待完善。为了得到合乎逻辑的理论，狄拉克设法建立一种对时间和空间坐标来说都是线性的微分方程，正是这种电子的波动方程导出了自旋与磁距的正确性。狄拉克的这一发现纯粹是基于对数学美的追求，事前丝毫未想到要给出电子的这种物理性质。

人的心灵深处感到美的东西，后来竟然在外部自然界得到了实现。这是在科学发现中最使人惊异的现象之一。美感在科学发现中之所以能起这么大的作用，首先是由于美和真之间存在微妙的统一关系。美的一些标志，很深刻地反映了未知真理的若干特征。科学家凭美的直觉有可能在理智抓住真理之前，先领悟了其中显现出的美，并以美为中介现实地获致真。一个由具有非常强的美学敏感性的科学家发展的理论，即使在它公布之初看起来不那么真，最后仍可能是真的。

美对于发现真的重要意义在一切时代都得到承认和重视。"美是真理的光辉" 这是一句拉丁格言，它告诉我们，"探索者最初是借助于这种光辉，借助于它的照耀来认识真理的。"

三、科学发展的推动力

数学符号美的魅力是科学发展的推动力。

数学符号的简单性是数学家追求的主要目标。拉丁格言"简单是真理的标志"以大字刻在哥廷根大学的物理学报告厅里，是作为对于那些将发现新事物的人们的一种告诫。所谓"简单性原理"就是科学家们有意识或潜意识所遵循的一种方法原则。历史表明，数学符号的发展总是遵循着"简单性原理"不断革新。例如：

$$\underbrace{a+a+\cdots+a}_{n\text{个}}\xrightarrow{\text{简化}}n\times a$$

$$\underbrace{a\times a\times\cdots\times a}_{n\text{个}}\xrightarrow{\text{简化}}a^n$$

$$\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}\longrightarrow f'(x_0)$$

$$\lim_{|\Delta x_i|\to 0}\sum_{i=1}^{n}f(\xi_i)\Delta x_i\longrightarrow\int_a^b f(x)dx$$

等等。

著名数学家傅里叶在创立"傅里叶级数"时，也进行了有关简单性的考虑，正如他自己所说："每一个数学函数，无论多复杂，总可以表示为某些简单的基本的函数（即相当于形成音乐中的纯音，或光学中的纯色的那种函数）之和。"傅里叶的这一成果不仅对后来关于偏微分方程积分的数理方法有启迪，而且对数学中函数概念的发展也起了促进作用。正如恩格斯所说："像黑格尔是一首辩证法的诗，傅里叶是一首数学的诗"。

二进制也可以说是从数字表达的简单性考虑所引出的产物，它推动了电子计算机的发展。这是计算数学的一场革命，它不仅对于数学发展来说开创了一个新的领域，而且对于整个自然科学的发展所产生的影响也十分深远。

数学中的各种最简形式（如最简多项式、最简分式、最简根式等）正是源出于对简单性的追求。

与最简形式相关，数学中的各种标准形式也是源出于对简单性的追求。

几乎可以说，数学家对于各种解析式、函数、方程等等符号总要设法寻求其最简形式、标准形式。这无疑是推动了数学发展。

对美的追求还推动了方法论的发展。所谓“补美法”就是最典型的事例。

对于美感与发现，创造心理学工作者致力于创造者的心理分析，揭示美感的作用；方法论学者则从中概括出补美原则。

什么叫补美法？美容师的工作就是帮人们补美。不过，科学中的补美是高层次的，要按照科学美的基本内容来补美。当一种理论尚未达到美的境界时，就必须继续进行创造、发展，“按照美的规律来创造”，这个过程中体现的方法，人们称之为“补美法”或“臻美法”。在第二段中，我们看到麦克斯韦实际上就是使用了补美法。

数学是自然科学的“皇后”，科学中的补美必须考虑到数学美，也就是要按照数学美的基本内容来补美。

下面我们根据椭圆的定义，建立椭圆的标准方程。这是高中学生所熟悉的一套，我们这里是为了阐明其中的补美手法。

平面上与两定点F_1、F_2的距离之和为常数(大于$|F_1F_2|$)

的点的轨迹（或集合）叫做椭圆。两定点叫做椭圆的焦点。

如图2－2，以过两焦点F_1和F_2的直线作为Ox轴，线段F_1F_2的中点O为原点，建立直角坐标系。设两焦点之间的距离为$2c(c>0)$，那末焦点F_1的坐标是$(c, 0)$，F_2的坐标是$(-c, 0)$。

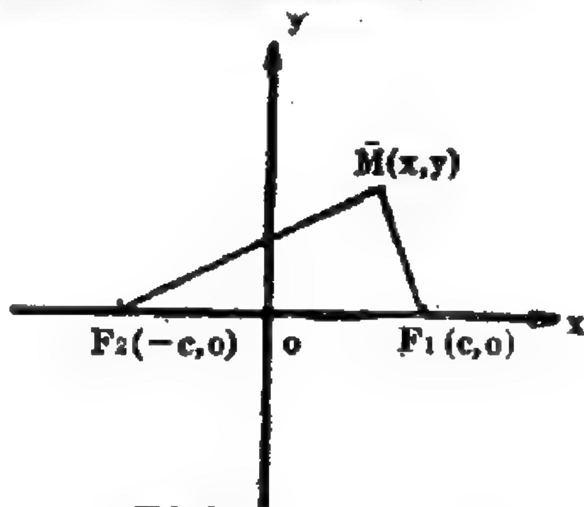


图2－2

设$M(x, y)$为椭圆上的任意一点，它到两焦点F_1及F_2距离之和用$2a(a>0)$表示，那末根据椭圆的定义，得

$$|MF_1|+|MF_2|=2a$$

即 $$\sqrt{(x-c)^2+y^2}+\sqrt{(x+c)^2+y^2}=2a \qquad (1)$$

照理方程(1)就可作为椭圆的方程，但因它不符合数学美的"简单性"要求，因此，必须化简。

方程(1)可化简为

$$(a^2-c^2)x^2+a^2y^2=a^2(a^2-c^2) \qquad (2)$$

用$a^2(a^2-c^2)$遍除上式的各项（不难看出，$a>c$，$a^2-c^2>0$），得

$$\frac{x^2}{a^2}+\frac{y^2}{a^2-c^2}=1 \qquad (3)$$

方程(3)比方程(1)简单多了，但是，它还不合数学美的要求。我们知道，椭圆具有对称性，那么，它的方程也应该具有对称性。对称性历来是数学家追求的目标。现代，考察对称性已成为数学研究中的重要美学指导思想。正如韦尔所说："对称性不

管你是按广义还是按狭义来定义，其含义总有一种多少时代以来人们试图用以领悟和创造秩序美和完善性的观念。”根据补美法，我们还要对方程(3)进一步加工，使y^2的分母与x^2的分母取得相同的形式，也就是说，要把a^2-c^2也写成一个正数的二次幂形式。

由于$a^2-c^2>0$，于是可令

$$a^2-c^2=b^2(b>0)$$

代入(3)式，得

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \qquad (a>b>0)$$

这个方程就是椭圆的标准方程。

可以证明，a正好是椭圆长半轴的长；b正好是椭圆短半轴的长。

我们从这个初等问题中，很清楚地看到了一个事实，根据数学美的要求用补美法引进的数“b”，有着鲜明的几何意义，而且果真符合“对称性”的要求！体现了美与真之间的统一性。正如著名数学家A·鲁滨逊所说：“纯粹数学的世界在很大程度上是由我们的关于数学美及纯粹数学的重要性的含糊的直觉来调整的。”

§5　某些符号刺激数学发展

有些数学符号、数学表达式具有神奇的诱惑力，为了对它们进行定性或定量的研究，几百年来，吸引了成千上万的学者为之奋斗，如此，不仅促进问题本身逐步解决，还推动了数学思想的发展，促进数学方法的革新。

姑且不说哥德巴赫猜想中的符号“$1+1=2$”是怎样令人着迷，这里只讲一个小小的字母符号“π”。

人们早已看出，圆的周长和直径的比是个与圆的大小无关的常数，并称之为圆周率。

1600年，英国的威廉·奥托兰特首先使用$\frac{\pi}{\delta}$表示圆周率，他的理由是，因为π是希腊文“圆周”一词的第一个字母，而δ是希腊文“直径”一词的第一个字母，根据圆周率的定义，理该应用$\frac{\pi}{\delta}$表示圆周率，但在推算圆周率的过程中，人们常用直径为1的圆，即令$\delta=1$，这样$\frac{\pi}{\delta}$就等于π了。1706年英国的琼斯首先提出改用π表示圆周率，1737年欧拉也在他的著作中用π表示圆周率。后来被数学家们广泛接受，一直沿用至今。

π是数学中的一个重要常数，在数学与自然科学中有广泛的应用。计算圆的周长和面积，计算球、球缺、球台、圆柱、圆锥、圆台等各种旋转体的表面积和体积等等都不可避免地要求助于π。伽里略在研究钟摆时，发现单摆周期T具有公式：

$$T=2\pi\sqrt{\frac{l}{g}}$$

库仑在研究两个带电质点的相互作用力时，得到公式：

$$F=\frac{1}{4\pi\varepsilon}\cdot\frac{Q_1Q_2}{r^2}$$

高斯在研究测量误差分布时，得到“正态分布”的密度函数：

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}$$

如此等等，举不胜举。难怪有人说π在数学中扮演着“不可一日无此君”的重要角色。

π值到底是多少呢？数学史表明，许多国家都十分重视圆

周率π 的计算。一位德国数学家评论道："历史上一个国家所算得的圆周率的准确程度，可以做为衡量这个国家当时数学发展水平的一个标志。"最初，数学家们希望能找到一个能准确表达π 值的分数，在各种尝试都失败了以后，人们发现这个分数是不存在的。1794年勒让德证明了π 是无理数，即不能用两个整数的比表示。1882年德国数学家林德曼证明了π 是超越数，即不可能是一个整系数代数方程的根。尽管如此，古今中外很多数学家都孜孜不倦地寻求过π值的计算方法。

首先从理论上给出π 值的正确求法的是公元前200 年 间的古希腊数学家阿基米德。他用圆外切与内接正多边形的周长从大、小两个方向上同时逐步逼近圆的周长，并用穷竭法（当时人们还未掌握极限法）巧妙地求得

$$3\frac{11}{71}<\pi<3\frac{1}{7}$$

公元前150年左右，另一位古希腊数学家托勒密用弦表法（以1°的圆心角所对的弦长乘以360再除以圆的直径）给出了π的近似值3.1416。

公元200年间，我国数学家刘徽用"割圆术"推算了π 的近似值。尽管结果没有超过托勒密，但是他在方法论上作出了划时代的成就，他第一次提供了求圆周率的科学方法——割圆术，体现了极限观点。刘徽的方法与阿基米德有所不同，他只取"内接"不取"外切"，利用圆面积不等式推出结果，起到事半功倍的效果。之后，祖冲之在圆周率计算上取得了世界领先地位，求得"约率"$\frac{22}{7}$和"密率"$\frac{355}{113}$（又称"祖率"），得到

$$3.1415926<\pi<3.1415927$$

这一惊人的记录，直到一千多年以后才被打破。可惜，祖冲之的计算方法后来失传了。人们推测他用了刘徽的割圆术，这仅是推测，究竟是什么方法？还是一个谜。

15世纪时，伊斯兰的数学家阿尔·卡西，通过分别计算圆内接和外切正 3×2^{28}边形周长，把π 值的计算推到小数点后16位，打破了保持上千年的祖冲之记录。

采用正多边形周长或面积的办法来求圆周长或圆面积，毕竟有很大的局限性，工作量过大。对圆周率π 值的计算的重大突破是寻求π 的解析表达式。1579年法国韦达发现了关系式

$$\frac{2}{\pi}=\sqrt{\frac{1}{2}}\cdot\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}\cdot\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}}\cdots$$

虽然这个公式在实际计算时会遇到多次开方的巨大困难，但它毕竟首次摆脱了计算π 离不开几何学的陈旧方法，开创了用解析的方法计算π值的新路子。

1650年瓦里斯把π表示成无穷乘积的形式

$$\frac{\pi}{2}=\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdot\frac{8}{7}\cdot\frac{8}{9}\cdots$$

稍后，莱布尼茨发现

$$\frac{\pi}{4}=1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\frac{1}{9}-\cdots$$

接着，欧拉证明了

$$\frac{\pi^2}{8}=1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\cdots$$

上述这些公式看来都不太复杂，但要用它们来计算π 值，工作量都很大。

1777年法国数学家蒲丰提出他的著名的投针问题。依靠它，可以用概率方法得到π 的近似值。假定在平面上画一组距离为a的平行线，向此平面任意投一长度为$l(l<a)$的针，若投针次

数为n，针与平行线中任一条相交的次数为k，则有

$$\pi=\frac{2ln}{ak}$$

很多人作过实验，1901年有人投针3408次得出$\pi=3.1415926$。如果把针长取为$l=\frac{a}{2}$，计算将简化为

$$\pi=\frac{n}{k}$$

用这种方法计算π 值，工作量不大。尽管在精确度上未突破前人记录，但是它具有方法论意义。从此，引进了一种方法：建立一个概率模型，它与某些我们感兴趣的量(这里是 常数π)有关，然后设计适当的随机试验，并通过这个试验的结果来确定这些量。

随着电子计算机的发展，现代，人们已按照上述思想建立起一类新的方法，即蒙特卡洛方法。

π 值的计算方法的最大突破是找到了它的反正切函数表达式。

1671年，苏格兰数学家格列哥里发现了无穷级数

$$\operatorname{arctg}x=x-\frac{x^3}{3}+\frac{x^5}{3}-\frac{x^7}{7}+\cdots\qquad -1<x\leqslant 1$$

1706年英国数学家麦欣首先发现

$$\pi=16\operatorname{arctg}\frac{1}{5}-4\operatorname{arctg}\frac{1}{239}$$

用它来计算π，只要按反正切函数的展开式计算$\operatorname{arctg}\frac{1}{5}$的前 6 项并计算$\operatorname{arctg}\frac{1}{239}$的前 2 项，我们就可在十几分钟里得到$\pi$ 的相当精确的近似值3.141592653，可见采用级数的方法，其速度远远胜过古典算法。

接着，勒让德、高斯等数学家相继发现了一系列有关π 与

反正切函数间的关系式，利用反正切函数的无穷级数展开，使得π值的计算迅速突破了100位大关。

本世纪50年代以后，圆周率π的计算开始借助于电子计算机，从而出现了新的突破。目前有人宣称已经把π计算到了一亿位甚至十亿位以上的有效数字。

对于实际应用来说，现代科学技术只要用到π的十位小数也就足够了。用祖冲之的密率计算赤道长度，误差仅3米左右，高精度的人造卫星的测量也不过如此。人们之所以还在借助电子计算机对π的准确值穷追不舍，并不完全是出于数学家们的兴趣。人们想借助这一典型的带挑战性的课题，不断提高计算机的硬件与软件功能，不断创造新的更强有力的计算方法。另外，还有更多的意义。一种理由是为了从统计上获悉π的各位数字是否具有某种规律性，例如，是否具有"正态性"。如果在其十进制展开式中所有数码以等频率出现的话，一个实数称为单纯正态的；如果所有同样长的数码均以等频率出现的话，它称为正态的。人们从1949年在$ENIAC$上计算π值以来，就想从统计上来探讨它是否具有这种性质。然而，现在尚不知道π是不是正态的，以至是不是单纯正态的。

竞争还在继续，正如有人所说，数学家探索中的进程也像π这个数一样：永不循环，无止无休……。以上仅是π的"马拉松计算"年表中的一部分。它已足够反映一个事实：π刺激着数学机体的多种"神经"，促进了数学机体发育。

§6 某些符号的"特异功能"

"在数学符号中好像存贮了一定量的智力，这种智力释放

出来时就能产生几乎是爆炸性的威力。这就像是强大的发动机，我们借助它竖起智力的结构；如果没有这种支持，我们的能力就无法进入这种结构。”

“有些数学符号似乎具备一种神奇的力量。”

甚至还有些数学家说，数学符号具有魔力、神功、异乎寻常的效力。

以上种种说法，其基本思想都是指数学符号，尤其是某些特殊的符号具有“特异功能”，推动了数学的发展。

这里也用一个小小的字母符号——e来阐明。

大家熟知，e是数列

$$\{(1+\frac{1}{n})^n\}$$

的极限，即

$$e=\lim_{n\to\infty}(1+\frac{1}{n})^n$$

然而，该数列是怎么发现的呢？这是众多读者感兴趣的问题。这问题与对数的发明有关。

苏格兰人纳皮尔约在1594年创立了对数，那时指数的概念尚未完成，也没有指数符号。纳皮尔花了20年的心血，编造了世界上第一个对数表，实际上是一个正弦对数表。纳皮尔的对数是相对于底数为

$$(1.0000001)^{10000000}$$

的对数。该底数可以写成

$$(1+\frac{1}{10000000})^{10000000}$$

的形式，可见纳皮尔实际上已经接触到数列

$$\{(1+\frac{1}{n})^n\}$$

把数列$(1+\frac{1}{n})^n$作为对数的底数，如果n越大，真数的间隔就越小，这样的对数表就越有用。人们很自然地想到一个问题：让n无限地增大，用它的极限

$$\lim_{n\to\infty}(1+\frac{1}{n})^n$$

作为对数的底不是更好吗？

数学家们证明了这个数列的极限的确是存在的。然而，它是一个无理数，人们不可能用数码写出它的准确数值，因此，只能引进字母符号来表示。1727年欧拉用"e"来表示这个极限，即

$$e=\lim_{n\to\infty}(1+\frac{1}{n})^n$$

符号e"出身非凡"。无论在理论上还是在实用上，e这个数都有特殊的重要地位。

以e为底的对数称为自然对数，通常不再把底数e写出，简记为$\ln x$（这相当于$\log_e x$）。

果然，e引入后，对于编制对数表有显著的效果。利用下面的级数可以很方便地编制出对数表。

由 $$\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots \quad -1<x\leqslant 1$$

$$\ln(1-x)=-x-\frac{x^2}{2}-\frac{x^3}{3}-\frac{x^4}{4}-\cdots \quad -1\leqslant x<1$$

得 $$\ln\frac{1+x}{1-x}=2\left(x+\frac{x^3}{3}+\frac{x^5}{5}+\cdots\right),\quad -1<x<1$$

设k是任意的自然数，命$x=\frac{1}{2k+1}<1$，则

$$\frac{1+x}{1-x}=\frac{k+1}{k}$$

$$\therefore \quad \ln\frac{k+1}{k}=2\left(\frac{1}{2k+1}+\frac{1}{3(2k+1)^3}+\frac{1}{5(2k+1)^5}+\cdots\right)$$

或
$$\ln(k+1)=\ln k+2\left(\frac{1}{2k+1}+\frac{1}{3(2k+1)^3}+\frac{1}{5(2k+1)^5}+\cdots\right)$$

根据这个公式，从$\ln 1=0$出发，就可以循环地计算出所有的自然数的自然对数。比如要想精确到小数点第6位，只要取括号内的前6项就行了。当然，考虑到循环计算有累计误差，实际上应多取几项。

可见，有了e之后，对数表的编造速度就大大加快了。如果要编造常用对数表，可以利用已造好的自然对数表，按换底公式算出相应的常用对数的值。

e的另一个显赫功劳是，它帮助人们证明了π的超越性，从而彻底解决了“化圆为方”问题。

数学史家曾指出，在整个数学史中，很难找出像三大尺规作图难题那样具有经久不息的魅力。二千多年来，无数的聪明才智曾倾注在这几个问题之中而得不到丝毫结果。实际上这三个问题都是不可能用尺规经有限次的作图步骤来解决的。1637年笛卡尔创建解析几何，尺规作图的可能性才有了准则。1837年凡齐尔（1814—1848年）证明了“三等分任意角”及“倍立方”都是不可能的，然而，关于“化圆为方”问题的解决却还有较大的麻烦。在“化圆为方”问题中涉及到$\sqrt{\pi}R$（R是给定圆的半径）这样的几何量，要解决“化圆为方”问题，就需要对π作定性的研究——π是不是超越数？

一个数，如果不是任何整系数代数方程的根，就叫做超越数。“超越数”这一名词是欧拉首先引入的，意思是这种数“超

越了代数方法的能力”。不过欧拉并未给出任何具体的超越数。直到一百多年后的1844年，法国数学家刘维尔才具体构造出一个超越数，从而证明了超越数确实存在。其实，现在人们都知道，超越数比代数数还要多得多。

人们早就猜测π和e可能是超越数，可是迟迟未得到证明。直到1873年，法国数学家埃米特才证明了e是超越数，至于π的超越性仍未解决。埃米特说："我不敢试着证明π的超越性。如果其他人承担这项工作，对于他的成功没有比我再高兴的人了。”不久，1882年，德国数学家林德曼在埃米特证明e是超越数的基础上，借助欧拉公式$e^{i\pi}+1=0$，终于证明了π也是超越数。

知道了π是超越数，“化圆为方”问题也就迎刃而解了。由解析几何可以知道，一个几何图形可以用尺规作出，必须并且只需其中所要作出的几何量可以由给定的几何量经过有限次的四则运算和开平方获得。因为π是超越数，不可能由R经有限次四则运算和开方而获得$\sqrt{\pi}R$，所以用尺规作图来“化圆为方”是不可能的。

这样，由于e的帮助，这个二千多年的悬案才宣告了结。

人们对于数e的真正认识，是在17世纪中叶，数学家们发现了双曲线下的面积与自然对数之间的关系

$$\int_0^x \frac{1}{x}dx=\ln x$$

之后。那时，人们逐渐了解到很多重要的函数、重要的极限、微分与积分都与数e有着极为密切的关系，自然对数的求导运算特别方便，所以在理论研究时宁愿采用自然对数而不用常用对数。另一方面，在很多具有复利率的现象中，人们研究量的

变化时，往往是由$A\cdot(1+\frac{n}{100})^{t}$型结构式，经过极限过程与数$e$相联系，因此很多自然规律都表现为以$e$为底的指数形式，例如

放射性元素铀的衰变规律为

$$N_t=N_0e^{-\lambda t}$$

大气压力P和高度h的关系为

$$P_h=P_0e^{-Lh}$$

电容器放电时电压V和时间t的关系为

$$V_t=V_0e^{-t/RC}$$

等等。

正因为这些原因，数e就如同π一样成了一个 特别 重要的无理数。几乎每一本微积分读本，哪怕最简单的读本，也要专门介绍数e，把$\lim\limits_{n\to\infty}(1+\frac{1}{n})^n$列为数学中最重要的基 本 极 限 之一。

科学史表明，数e的“特异功能”不仅推动数学 发展而且影响到其它科学技术的发展。

众所周知，发射人造卫星或宇宙飞船，必须使用火箭。在这问题中，e果真“就像是强大的发动机”有“爆炸性的威力”。

俄国的星际航行专家早就发现，火箭的运动速度可以用方程

$$V=w\ln Z$$

表示，其中V是火箭的理想速度，w是火箭发动机的 排气速度，Z是火箭的最初质量与发动机停止时的质量之比（简称为质量比）。

火箭的最终速度至少要达到第一宇宙速度（约8千米/秒）

才能将人造卫星送至运行轨道。

上述方程表明，火箭的运动速度V，决定于喷气速度w和质量比Z。

目前，最好的混合液体燃料也只能使喷气速度达到4千米/秒。同时，质量比也无法提高很多，最大不会超过10。如果按$w=4$千米/秒，$Z=100$的理想情况计算，火箭的理论速度可以达到9千米/秒，但除去重力和空气阻力等造成的损失(约为2千米/秒)，实际速度只有7千米/秒，还是无法将人造卫星送至运行轨道。

科学家们通过研究火箭运动方程，找到了提高火箭速度的办法，即采用多级火箭。由接连的几个火箭发动机组成一个动力系统推动火箭前进：第一级发动机使火箭达到一定的速度，燃料耗尽以后就自动脱落；第二级发动机立即接着工作，使火箭达到更高的速度，燃料耗尽以后又自动脱落；第三级发动机再接着工作……。

假设各级发动机的喷气速度和质量比都相同，即

$$w_1=w_2=\cdots\cdots=w_n=w,$$
$$Z_1=Z_2=\cdots\cdots=Z_n=Z,$$

则火箭的最终速度

$$\begin{aligned}v&=v_1+v_2+\cdots\cdots+v_n\\&=w_1\ln Z_1+w_2\ln Z_2+\cdots+w_n\ln Z_n\\&=nw\ln Z\end{aligned}$$

也就是在这种情况下，每增加一级火箭，就可使火箭速度提高一倍。

现代的宇航技术，完全可以使发动机的喷气速度达到$w=3$千米/秒，质量比达到$Z=3$。如此，只需三级火箭，就

可以使最终的理论速度达到

$$V=3w\ln Z=3\times 3\ln 3\approx 10(\text{千米/秒})$$

即使实际损失 2 千米/秒，也足以达到 8 千米/秒的第一宇宙速度。

可见，数e帮助人们揭示火箭运动规律，启发人们设计出多级宇航火箭，终于实现了古代神话中遨游太空的理想。

数学家O.G苏顿（O.G.sutton）写道："有这样一个奇怪的想法，数学所使用的少数符号，没有什么意义的纸上记号，居然对生活的模样做出了世人所知的如此多的贡献。如果一个中世纪学者现在又醒来，他将会认为这些符号是符咒组成的魔力公式，如果念得对，就可以给人们以战胜自然的能量。"

显然，这种奇怪的想法并不奇怪，它是客观现实的反映。

第三章　数学符号的思维功能

数学符号的动力作用是由数学符号内在的思维功能所决定的，动力作用是思维功能的外显形式。

符号和数学思维有密切的关系。数学符号是数学抽象思维的产物，数学的符号语言有助于思维。正如莱布尼茨所说："记号是为了方便于发现，它多半是在记号简洁地去表示并能反映事物的内在本质的时候。这时思维活动以惊人的方式得到简化……"。如果说，数学是思维的体操，那么，数学符号的组合谱成了"体操进行曲"。

数学家苏顿（O.G.Sutton）曾经提出一个问题：

"我们是根据我们自己规定的规则创造了符号，那么在这种符号上的运算操作又怎样能揭示出什么超出我们感觉的东西呢？换句话问就是，我们怎能用数学去发现任何东西呢？"

苏顿接着说："上面的提问都不像是我们这一代人能够给出满意答案的问题。我们至多只能指出这个答案的几条线索。"

徐利治先生也提出过类似的问题：为什么某些从不同角度确定的符号竟然会统一在一个极简单的关系式中呢？这种关系是否与人的大脑的数学思维机能和特性有关呢？他说，这显然是研究数学与思维关系的最诱人的问题之一。

自从1980年钱学森教授倡导建立思维科学以来，研究思维的规律和方法，已成为热门课题。研究数学符号的思维功能是

揭示大脑的数学思维机能和特性的需要，这一章试图给出“几条线索”。

§1 思维活动的物质载体

数学符号是交流与传播数学思想的媒介，可用以传递大块的信息。

数学符号是数学抽象物的表现形式，它有自己的思想内容。数学符号按一定规则组织起来，就成为数学思维活动的物质载体。人们通过用数学符号组成的语言交流数学思想，认识数学世界的奥秘，并把数学成果应用于人类各种实际问题。

数学符号的载体功能大致表现于下面三方面：

一、表示一般规律

高度的抽象性，保证了应用的广泛性。数学符号是抽象思维的产物，它可用以表示一般的数量关系及其变化规律。

先看数字符号。

符号“0”看起来是多么简单的符号，但却有极其丰富的内容。

在实数系、有理数系和整数系里，它均表示唯一的中性数。

“我有0元钱”其中“0”表示“没有”。

“今天天气温度是0度”，其中“0”表示了这一天的温度的高低情况，已不是“没有”温度的意思。

在集合代数中，“0”表示“空集”。

在开关代数中，“0”表示“关”。

在命题代数中，“0”表示“假”。

此外，“0”还可表示零点、零元、零向量、零函数、零环、零代数……

在科学研究中，“宇宙之大、粒子之微、火箭之速”如果只靠自然语言是难以表达的，必须依赖于数字符号。比如，光的速度的数字表示是3×10^8米/秒，天文学中的距离单位——光年的数字表示是9.46×10^{15}米；而在微观世界中，核子（如质子、中子等）的直径约为3×10^{-15}米。

再看变元。

变元可以代替一类符号中的任意一个符号；字母x、y等可作为任意集合元素的变元。以二次函数$y=ax^2$为例，它可用以描述自由落体的运动规律 $S=\frac{1}{2}gt^2$，可以刻划动能与速度、质量之间的关系 $E=\frac{1}{2}mV^2$，也可用以表达圆的面积$S=\pi r^2$。它可以表示一个二次曲线；表示某天体运动的轨迹；也可以表示某圆锥面的截线。

要研究自然的奥秘，进行抽象思维，揭示客观规律，必须使用比自然语言远为优越的形式化的符号语言。伽利略说宇宙这部书是用数学语言写成的，这是有其深刻道理的。从质子、中子、电子这些基本粒子的运动，到月亮、地球、太阳这些天体之间的关系，从简单的机械运动，到复杂的化学变化，从逻辑思维的基本规律，到不容辩驳的推理和证明，无不可以用数学语言清晰、准确地描述出来。

例如，牛顿力学的惯性定律、作用力与反作用力定律、万有引力定律可以用数学表达式分别表示为

$$\vec{F}=m\vec{a}$$

$$\vec{F}_1 = -\vec{F}_2$$

$$F = G \cdot \frac{m_1 \cdot m_2}{r^2}$$

以此为基础，既可以精确地描述太阳系中的行星、卫星和彗星的运动，又可以完满地解释地球上潮汐和其它物体的运动。

当然，最典型、最突出的例子应推第二章§4所提到的爱因斯坦的质能关系式：$E=mc^2$。该公式只用极少的几个字母，却深刻地揭示了微观、宏观、宇宙无数质能变化现象的规律。特别地，它揭开了原子内部的秘密，使人们发现了蕴藏在原子核内部的巨大能量，帮助人们发掘出取之不尽，用之不竭的能源——原子能。

用华罗庚的话来说，“化工之巧、地球之变、生物之谜、日常之繁”均可用数学符号来描述。随着数学符号运用范围的日益扩大，许多科学家干脆就把数学的符号语言称为“科学的语言”。

二、构造数学模型

数学符号可用以构造数学模型。

“只有在详尽给出现实世界的一个模型之后，数学才能出场。数学的每一个应用都依赖于模型。”数学应用于实际的关键在于用数学语言描述出所要研究的问题，使之构成一个数学问题，这个数学问题被称为研究对象的数学模型。现代，数学模型方法不仅是处理数学理论问题的一种经典方法，而且也是处理科技领域中各种实际问题的一般数学方法。

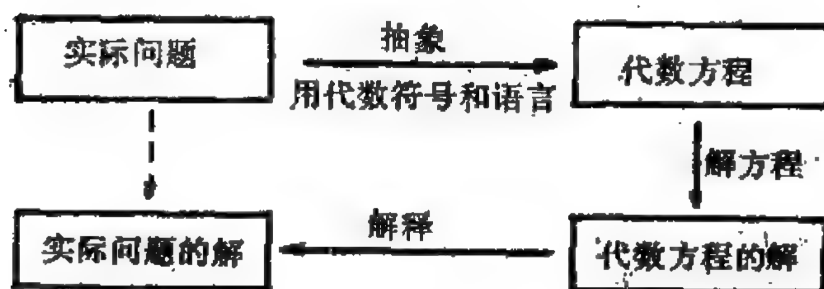
笛卡尔在他的《思维的法则》一书中，曾设计了一种“能解各种问题”的“万能代数模型”：

首先，把任何问题化为数学问题；

其次，把任何数学问题化为一个代数问题；

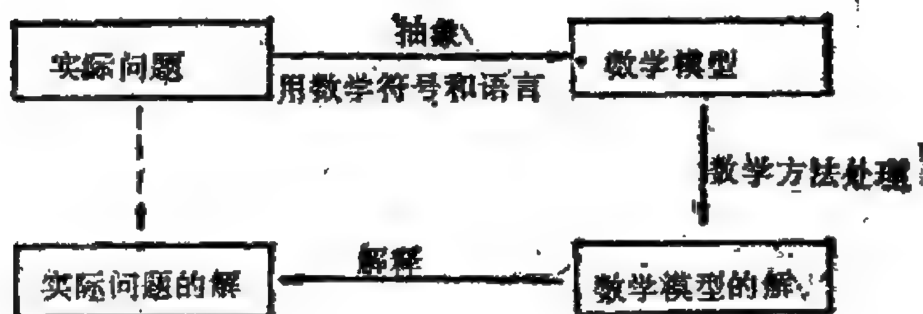
最后，把任何代数问题归结到去解一个方程式。

"万能代数模型"的框图为：



显然，这种"万能方法"是有漏洞的，在某些情况下是不适用的。笛卡尔未能完成他的"法则"，然而，他的想法有着某些深刻的道理，仍不失为"一个伟大的设计"，即使失败了，它对数学方法论的发展的影响是不可抹杀的。虽然笛卡尔的"万能方法"不能应用于所有的情况，但是仍能适用于非常多的情况，其中包括一些十分重要的情况。用"列方程"的方法解应用题时，正是按照笛卡尔的基本思想去做的。布列方程就是建立数学模型。

现代数学模型方法正是笛卡尔"万能代数模型"方案的继续延伸和直接推广，其框图为：



可以看出，制作数学模型，离不开数学符号。正如数学家M·黎伯特（M.Report）所说："数学作为创造性的学科，按三个基本步骤进行：

（1）体验一个问题，并从中发现一个模式；

（2）定义一个符号系统来表达这个模式；

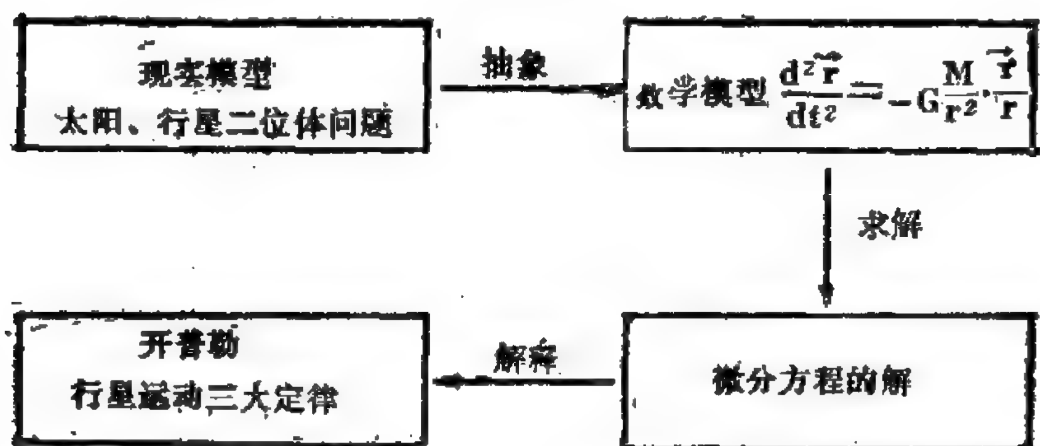
（3）把这个符号系统组织为一个系统的语言。"

广义地讲，数学中各种基本概念（如实数、向量、集合、群、环、域、范畴、线性空间、拓扑空间…）以及各种方程式（代数方程、函数方程、微分方程、差分方程、积分方程）等都可称作数学模型。因为它们都是以各自相应的现实原型作为背景而加以抽象出来的数学对象。

狭义地讲，只有那些反映特定的问题或具体事物系统的数学关系结构才叫数学模型。在应用数学中，构造数学模型的目的都是为了解决具体实际问题。

17世纪时，伟大的数学家、物理学家牛顿研究行星绕太阳运动的规律（太阳、行星二位体问题），就是运用了数学模型方法。他忽略了其它各种天体的影响，并把太阳与行星都看成是质量集中在中心的理想质点，从而根据万有引力定律得出了行星绕太阳运动的微分方程，再通过微分方程的解，推导出了

开普勒的行星运动三大定律。即：



科学理论的数学化已经成了现代科学发展的基本特点之一，而科学理论数学化的关键就在于建立适当的数学模型。例如，在生物学中，可以用微分方程模型（$dN=kNdt$）来描述细胞的繁殖和人的增长；可以用伯努利二项分布模型（$P\{X=k\}=c_n^k p^k q^{n-k}$）来描述生物群落在某个区域内的分布；可以用线性代数网络模型来描述生物神经系统中信号的传递等等。这样，就使得对这些生物现象的研究上升到了精确可靠的定量分析的水平。故有人说，这种借助数学模型来显示生物现象变化规律的方法，使生物学获得了突飞猛进的第二次生命。

现代数学模型已经非常广泛地应用于自然科学、工程技术、经济科学、军事科学、交通运输等科技领域。在它的各种应用中，作为数学模型的“数学关系结构”都应该是借助于数学概念和符号刻划出来的某种结构。

三、表达数学思维模式

数学符号可用以表达数学思维模式。

数学思维模式是人脑对数学信息进行加工的相对稳定的思维样式。数学思维模式与数学模型有区别又有关联，它是由一定的数学知识结构与数学思维方式结合而成的动态系统。

怀特海曾说过："数学是对模型的研究"，整个数学实质上是由各种层次的大大小小的各种模型所组成的模型系统。而数学思维模式就是识别数学知识模型的有效途径和形式化、标准化的思维程序。因此数学思维模式具有约简思维过程，降低思维强度，提高思维效率的认识功能。

数学思想、观点是数学知识内容和数学思维规律的高度概括，适用范围较广，因而是较高层次的或一般的数学思维模式，例如，交轨模式、方程模式、递归模式、叠加模式、逼近模式和映射模式等等。数学中的基本原理以及某些典型的数学问题的解法是数学思维过程中的思维反应块，相当于房屋建筑中的一些组合构件，它们适用于某一类特定问题的化归，因而，是一些较低层次的具体的数学思维模式。例如，勾股定理，韦达定理等都是具体的数学思维模式。

为了形成那些具体的数学思维模式，数学符号往往是不可缺少的。

众所周知，整数的加法和乘法服从五个算术基本规律——加法和乘法的交换律、结合律、乘法关于加法的分配律。这些运算规律，是人们千万次运算经验的总结，为了要表达这些具有普遍性的规律，人们不能用诸如1、2、3这种表示特定数的符号，这时，可用变元来作为表示整数的符号，于是，五个算术基本规律可叙述为

$$a+b=b+a$$

$$ab=ba$$

$$a+(b+c)=(a+b)+c$$
$$(ab)c=a(bc)$$
$$a(b+c)=ab+ac$$

这组公式堪称算术、代数乃至整个数学的基础。它借助变元，把人们的运算经验表示为“相对稳定的思维样式”。正如日本学者池上嘉彦所说：“使一般性的理解方法固定下来的是符号”。人类使用符号表达数学思维模式，“肯定和维持已经诞生的人类文化秩序，使之功能化。”

关于极限的定义，也是很典型的例子。极限的思想可追溯到古代，但是直到牛顿时代，人们还未建立严格的极限定义。那时牛顿所运用的极限概念还只是直观性的语言描述（以数列为例）：

对于数列$\{a_n\}$，如果当n无限增大时，a_n无限地接近于常数A，那么就说$\{a_n\}$以A为极限。

这种描述性语言，人们容易理解，但它没有定量地给出两个“无限过程”之间的联系，不能作为科学论证的逻辑基础。正因为当时缺少严格的极限定义，微积分理论才受到人们的怀疑、攻击。如在瞬时速度概念中，究竟Δt等于零还是不等于零？牛顿自己也左右为难，无法解释。众所周知，这个微积分理论基础问题，后来是柯西与维尔斯特拉斯解决的。他们建立了严格的极限定义（仍以数列为例）：

设$\{a_n\}$为一数列，A是一常数，如果对任何$\varepsilon>0$，总存在自然数N，使得当$n>N$时，不等式$|a_n-A|<\varepsilon$恒成立，则称$\{a_n\}$的极限是A。

这个定义，借助不等式符号，通过变元ε和N之间的关系，定量地、具体地刻划了两个“无限过程”之间的联系。因此，

这样的定义是严格的，可以作为科学论证的基础，至今仍不显得陈旧。

这种符号语言不仅给出数列极限的严密定义，而且给出了证明数列极限的思维模式——一种“形式化、标准化的思维程序”，要证明$\lim_{n\to\infty}a_n=A$，只要先任意给定一个小正数ε，再设法探求相应的“N”，使得$n>N$时，$|a_n-A|<\varepsilon$恒成立。

可见，数学符号具有表达数学思维模式的载体功能。正如池上嘉彦所说，符号的组合“能够规定思考方式”。

§2 符号暗示信息

符号具有意指作用，因此，能暗示信息。信号弹上天，暗示战斗开始；公路旁的急转弯警告标志，暗示驾驶员要减速。数学符号所暗示的信息量更大，可信度更高。

数字规律，字母符号或表达式的结构、特征等都有自己的思想内容，都能暗示某种信息。对这种信息的捕捉能力是科研工作者的需要，一个富有直觉洞察力的数学家，对这种信息的反应总是很敏感的。

数学特别是数论中好多重要定理都是从发现某种数学规律开始的，正如欧拉所说：“今天人们所知道的数的性质，几乎都是由观察发现的，并且早在用严格论证确定其真实性之前就被发现了。甚至到现在还有许多关于数的性质是我们所熟悉而不能证明的，只有观察才使我们知道这些性质。”这里，观察的作用就是发现数字间的各种规律。

欧拉在一本简短笔记中指出，对正值参数n，级数

$$1-\frac{x^2}{n(n+1)}+\frac{x^4}{n(n+1)(n+2)(n+3)}$$
$$-\frac{x^6}{n(n+1)\cdots(n+5)}+\cdots \quad (1)$$

对所有x值都收敛。并提出一个令人惊异的猜想：

由级数(1)所定义的函数当$0<n\leqslant3$时仅有实零点，且有无穷多个，但当$n>3$时，没有实零点。这里，他把n当作连续变动参数。

欧拉观察了关于$n=1, 2, 3, 4$，级数(1)的和及其零点：

n	和	零点
1	$\cos x$	$\pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2}, \cdots$
2	$\frac{\sin x}{x}$	$\pm\pi, \pm2\pi, \pm3\pi, \cdots$
3	$\frac{2(1-\cos x)}{x^2}$	$\pm2\pi, \pm4\pi, \pm6\pi, \cdots$
4	$\frac{6(x-\sin x)}{x^3}$	无实零点

欧拉观察出一个差异：在前三个实例中所有零点都是实的，在最后一个实例中没有一个零点是实的。欧拉注意到在前两个与第三个实例之间又存在颇为微妙的差异：对$n=1$与$n=2$，两个相邻的零点之间的距离是π（当$n=2$时，相邻于原点的零点例外），但对$n=3$，相邻零点间的距离是2π（相邻于原点的零点也例外）。这激起他进一步的观察：$n=3$时，所有零点都是二重零点。这是富有暗示性的结果。欧拉说，然而我们由分析得知"一个方程的两个根总是在由实根到虚根的过渡中相合。那么，我们可以弄懂为什么当n超过3时，所有零点突

然变成复的。”这些观察数据，对欧拉起了暗示作用，他从中发现：级数(1)的和函数当$0<n\leqslant3$时仅有实零点，且有无穷多个，但当$n>3$时，没有实零点。

在欧拉时代，超越方程零点的实性问题完全是新的，甚至在今天人们也不具备系统的方法来解决这样的问题。事实上，现今人们还不能证实黎曼的著名假设是真是假。因此，欧拉的猜想极其大胆、令人惊叹的。欧拉是从几个零星数据的特征去猜测整体的，表现出超人的直觉洞察力。读者也许会问，为什么欧拉只考虑$n=1$，2，3，4这几个实例就能进行推测。这是因为欧拉抓住了有力的旁证材料，获得关键的信息。

举例来说吧。大家熟知，带有参数k的方程

$$x^2-k=0$$

当$k>0$时有两个实根；当$k=0$时出现二重根；当$k<0$时有两个虚根。这就是说，“方程的两个根总是在由实根到虚根的过渡中相合”。也许欧拉的记忆中也有这样的潜知。为此，欧拉根据$n=1$，2，3，4时的观察资料，就有足够的理由推测：级数(1)的和函数当$0<n\leqslant3$时仅有实零点；当$n>3$时，没有实零点。这里$n=3$时的实例是欧拉作出猜想的最强有力的、最有启发性的旁证材料。

欧拉并未证实他自己的猜想，而只是对一些特殊的n的数值(如$n=\frac{1}{2}$，$\frac{1}{3}$，$\frac{1}{4}$)，进行了检验，经过多次核验，他相信其正确性。

时隔150年，G·波里亚证实了欧拉的猜想。详见G·波里亚的论文《关于欧拉论文的超越方程》(《意大利数学家学会会报》第5卷，1926，64—68页)。

符号的暗示作用对于数学解题也是可贵的。在数学解题中，正确思路的选择是头等重要的。正如G·波里亚所说：“解题的成功要靠正确思路的选择，要靠从可以接近它的方向去攻击堡垒。”就像公路旁的指示标志能暗示方向一样，数学符号能暗示解题思路。

运算符号蕴含了相应运算的基本性质，它所暗示的解题思路对于数学解题是头等重要的。

例 已知$y=3x^2+\sqrt{\arcsin x-\frac{\pi}{2}}+x-2$，求$\log_4 y$的值。

如果不假思索，直接进行对数运算是无效果的，为了解题的成功，必须选择正确的思路。仔细观察题意，可以发现式中有个算术根符号“$\sqrt{\quad}$”。符号是意愿的标志，传播一定的意义。这里，符号“$\sqrt{\quad}$”向我们传播一种信息——被开方数不小于零，暗示了题设中的隐含条件：$\arcsin x-\frac{\pi}{2}\geqslant 0$。再考虑到反正弦函数的值域：$|\arcsin x|\leqslant\frac{\pi}{2}$，本题就不难解决了。

对数运算的符号所传播的信息量更大。

例 求解关于x的方程

$$\log_{\left(cx+\frac{d}{x}\right)}x=-1 \quad (\text{其中}c,\ d,\ x\in R,\ c\neq 0)。$$

对数符号“log”本身，向我们传播了几个信息：$x>0$，$cx+\frac{d}{x}>0$，$cx+\frac{d}{x}\neq 1$。

再考虑方程中的等号，就可列出下列方程：

$$\begin{cases} x>0 & (1) \\ cx+\dfrac{d}{x}>0 & (2) \\ cx+\dfrac{d}{x}\neq 1 & (3) \\ x(cx+\dfrac{d}{x})=1 & (4) \end{cases}$$

考虑到(2)被包含在(1),(4)中；根据(4),(3)与$x\neq 1$等价,所以上述方程组又可简化为

$$\begin{cases} x>0 \\ x\neq 1 \\ x(cx+\dfrac{d}{x})=1 \end{cases}$$

进而解得

$$\begin{cases} x=\sqrt{\dfrac{1-d}{c}} \\ \dfrac{1-d}{c}>0 \\ \dfrac{1-d}{c}\neq 1 \end{cases}$$

于是，当$\dfrac{1-d}{c}>0$且$\dfrac{1-d}{c}\neq 1$时，方程有唯一解

$$x=\sqrt{\frac{1-d}{c}}$$

当$\dfrac{1-d}{c}\leqslant 0$或$\dfrac{1-d}{c}=1$时，方程无解。

数字符号的规律也能暗示解题思路。

例 设P为等边三角形ABC外一点，且$PA=3$，$PB=4$，$PC=5$，求$\triangle ABC$的边长。

分析 PA、PB已知，欲求AB的值，可在$\triangle PBA$内运用余弦定理，为此，需求$\angle APB$的度数。题中数字“3、4、5”有明显的特征，它暗示我们一个信息：有直角三角形或者可构造一个直角三角形以便使用。当然，所构造的直角三角形，必须对求$\angle APB$有利。为此，设法变更线段PC和PA的位置，使之和PB组成直角三角形。考虑到$\triangle ABC$是等边三角形，以点A为旋转中心，将$\triangle APC$逆时针旋转60°，则顶点P到达点M的位置，顶点C到达点B的位置（见图3－1）。PC到达MB的位置。

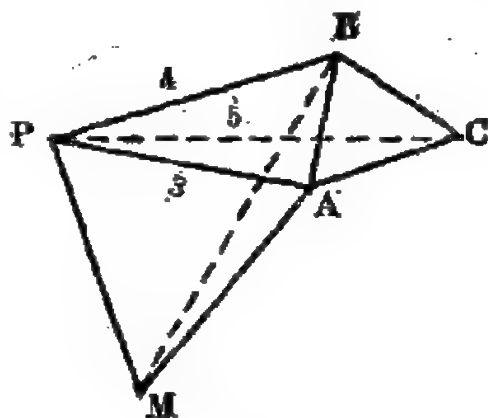


图3－1

$\because \quad \triangle PAC \cong \triangle MAB$

$\therefore \quad PA=MA$

$\quad\quad \angle PAM=60°$

$\because \quad \left\{\begin{array}{l}\angle MPA=60° \\ PM=PA=3\end{array}\right. \Longrightarrow \triangle MPB$为直角三角形

$\therefore \quad \angle APB=30°$

难点宣告解决。

在数学题题设条件中，字母符号所呈现出的各种特征，都暗示着解题思路。特别地，在题设条件里地位相同的未知量暗示着它们在解答中的地位也相同。根据这个原理在很多时候能使我

们预测到问题的解，或者发现解题途径。当然，其理由是不充足的。人们把这样的原理称为“不充足理由律”。

例 已知三正数a、b、c满足$a+b+c=1$，求证

$$(a+\frac{1}{a})^2+(b+\frac{1}{b})^2+(c+\frac{1}{c})^2\geqslant\frac{100}{3} \qquad (1)$$

分析 在题目的条件与结论中，a、b、c的“地位”是平等的，根据不充足理由律，我们可以预测当左边各项等于$\frac{100}{9}$，即$a=b=c=\frac{1}{3}$时，结论中的等号成立。依此猜测先将所需结论转化，再进一步探索证明途径。

$$\because\ (a+\frac{1}{a})^2+(b+\frac{1}{b})^2+(c+\frac{1}{c})^2$$

$$=(a^2+b^2+c^2)+(\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2})+6 \qquad (2)$$

令$A=a^2+b^2+c^2$，$B=\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}$，

当$a=b=c=\frac{1}{3}$时，$A=\frac{1}{3}$，$B=27$，而$\frac{1}{3}+27+6=\frac{100}{3}$。

因此，把(1)，(2)右边相比可以看出，要证明原式只需证明$A\geqslant\frac{1}{3}$，$B\geqslant27$。

先证$A\geqslant\frac{1}{3}$，即$3A\geqslant1$。

根据已知条件应该证明$3(a^2+b^2+c^2)\geqslant(a+b+c)^2$，这只要由$2(a^2+b^2+c^2)\geqslant2(ab+bc+ca)$两边各加$a^2+b^2+c^2$，即得。

再证 $B\geqslant27$。

$$B=\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}\geqslant3\cdot\sqrt[3]{(\frac{1}{abc})^2},$$

由 $a+b+c\geqslant 3\sqrt[3]{abc}>0$，

得 $abc\leqslant\frac{1}{27}$，$\frac{1}{abc}\geqslant 27$。

故得 $B\geqslant 3\sqrt[3]{27^2}=27$。

"不充足理由律"特别适用于对称性，数据和条件里的对称性往往在解里得到反映。在某种程度上，"数据和条件里的对称性"不仅仅被"求解对象"所反映，而且亦为"求解过程"所反映。人们把这种原理叫做对称性原理。

例 已知四面体$P—ABC$的六条棱长之和为l，并且$\angle APB=\angle BPC=\angle CPA=90°$，试求它的最大体积（美国第五届中学数学奥林匹克竞赛试题）。

解 设$PA=a$，$PB=b$，$PC=c$。

则 $AB=\sqrt{a^2+b^2}$，$BC=\sqrt{b^2+c^2}$，

$CA=\sqrt{c^2+a^2}$

根据题意，有

$$l=a+b+c+\sqrt{a^2+b^2}+\sqrt{b^2+c^2}+\sqrt{c^2+a^2} \quad (1)$$

$$V=\frac{1}{6}a\cdot b\cdot c \quad (2)$$

不难看出，(1)、(2)两式都是a、b、c的轮换对称式。这种对称性，使得a、b、c在求体积最大这个问题上所起的作用是相同的。根据不充足理由律，我们可以预测，当$a=b=c$时，体积最大。

我们由(1)、(2)先求出当$a=b=c$时四面体的体积。这时，

$$a=b=c=\frac{l}{3(1+\sqrt{2})}$$

$$V=\frac{1}{6}\left[\frac{l}{3(1+\sqrt{2})}\right]^3$$

下面只要证得

$$\frac{1}{6}abc \leqslant \frac{1}{6}\left[\frac{l}{3(1+\sqrt{2})}\right]^3$$

就证实了我们的猜想。

要证 $\frac{1}{6}abc \leqslant \frac{1}{6}\left[\frac{l}{3(1+\sqrt{2})}\right]^3$

只要证

$$(1+\sqrt{2})\cdot\sqrt[3]{abc}$$
$$\leqslant\frac{1}{3}(a+b+c+\sqrt{a^2+b^2}+\sqrt{b^2+c^2}+\sqrt{c^2+a^2})$$

由于 $\sqrt[3]{abc}\leqslant\frac{1}{3}(a+b+c)$，因而只要证

$$\sqrt{2}\cdot\sqrt[3]{abc}\leqslant\frac{1}{3}(\sqrt{a^2+b^2}+\sqrt{b^2+c^2}+\sqrt{c^2+a^2})$$

上式的确是成立的，这是因为

$$\frac{1}{3}(\sqrt{a^2+b^2}+\sqrt{b^2+c^2}+\sqrt{c^2+a^2})$$
$$\geqslant\sqrt[3]{\sqrt{a^2+b^2}\cdot\sqrt{b^2+c^2}\cdot\sqrt{c^2+a^2}}$$
$$\geqslant\sqrt[6]{2ab\cdot 2bc\cdot 2ca}=\sqrt{2}\sqrt[3]{abc}$$

现代，对称性原理已经渗透到自然科学的很多领域。杨振宁博士说过："三十多年来，我进行的物理研究工作，都同对称性原理和统计物理两大题目有关。"1979年，他在关于纪念爱因斯坦诞辰一百周年的会上，讲演的一个主题是"对称性决定相互作用"。

人们评价说，对称性原理"支配着理论物理学家创造了数学表达形式，它使我们懂得应该怎样创立理论，才能精确地描述自然界。"

§3 约简思维，促进思维“机械化”

数学家怀特海（A.H.Whitehead）曾说过：“一个良好的标记法，使头脑摆脱了不必要工作的负担和约束，使它集中于先进的问题，这就在事实上增加了人类头脑的能力。”在数学中，问题的陈述、推理的过程以及定量计算，都运用简明的数学符号，大大简化和加速思维的进程，促进思维“机械化”，思维的经济原则在数学中得到了高度的发挥。正如数学家马赫所说：“数学的力量在于它避免了一切不必要的思想而采取了最为经济的思维方式。序号被称之为数，这就已经形成了一种奇妙、简单而又经济的系统。当我们进行数的乘法运算时，对于乘法表的利用，就使我们能以利用先前已完成的结果，而不必每一次都去作重复的运算。又如当我们利用对数表时，同样也是利用已完成的计算去取代新的数值计算。再例如，当我们利用行列式去解方程组时，以及当我们把新的积分表达式分解成其它已知表达式的时候，我们就可看到拉格朗日或柯西的智力活动，他们总是以军事司令官的敏锐识别力，统帅着所有已完成的运算的‘军队’，并由此而去执行新的运算任务。”有些数学家还把数学说成是“极端的思维经济学”。

一、使用符号可以约简推理步骤

数学离不开推理。怀特海曾说过：“在最广泛的意义上说，数学乃是多种形式的必然的演绎推理的展开。”数学被称为演绎的科学，它的每一学科都要求从最少的几个原始概念和最少的几条公理出发，经过推理得到该学科的全部知识。例如，通常的

几何，是仅由五条公理（欧几里德公理）出发推导出来的，数论也是仅由五条公理（皮亚诺公理）出发推导出来的。如此等等，不胜枚举。罗素和怀特海在《数学的原理》一书中还说："纯粹数学是所有形如'p蕴涵q'的所有命题类。"其实，这种说法是片面的，它只反映了数学的一个侧面："已严格地提出来的数学是一门系统的演绎科学"。然而，"正在形成过程中的数学是一门实验性的归纳科学。"数学中还经常运用归纳推理、类比推理等思维形式。

符号与推理有密切关系。古代，斯多葛派的哲学家们提出符号问题，就是为了建立严密的三段论逻辑的需要。他们认为，符号间的关系就像是三段论向外部物质的投影。数学符号对于数学推理"看来是必不可少的"，使用合宜的符号系统可以使数学推理步骤变得简单，达到以最少的思维获得最佳效果的效能。

波里亚曾说过："对数学符号的重要性我们几乎总是不会估计过高的。现在使用十进制符号的计算工作者比古代不能以如此方便的形式记数的计算工作者要沾光得多。"所谓沾光，就是约简了思维。拉普拉斯也曾盛赞阿拉伯数字符号："用不多的记号来表示全部的数的思想，赋予它的除了形式上的意义外，还有位置上的意义，它之所以如此绝妙，正是由于这种简易（思维的方便）难以估量。"阿基米德堪称古今中外最伟大的一个数学家，高斯最敬佩阿基米德，可是高斯对阿基米德未能发明十进位制及其表示法表示遗憾，高斯说："令人不解的是，他怎么会没有看出这一点，假如阿基米德能作出这个发现的话，那么现在的科学该处在多么高的水平上呀！"可见，高斯对十进制符号的思维功能有极高的评价。

在古代，关于一元二次方程的根的一般讨论要写上200页

纸，而在现在只要1页纸就足够了。这是因为借助于代数的符号系统。当初，笛卡尔正是由于看到了符号代数有约简思维、加速进程的巨大威力，才着手把代数用于几何的伟大工作。之后，解析几何的发展，果真达到了笛卡尔的目标。不妨，看个实例：

如图3-2，过一圆的AB弦的中点M，引任意二弦CD和EF，连接CF和ED交AB弦于P、Q，

求证$PM=MQ$。

这是历史名题，因其图形奇巧，酷似翩翩起舞的蝴蝶，故名蝴蝶定理。它在1815年就被英国数学家霍纳(Horner)证明了，但至今仍引吸着众多的学者，证明方法很多，长短不一、难易不同。下面看看它的解析证明法。

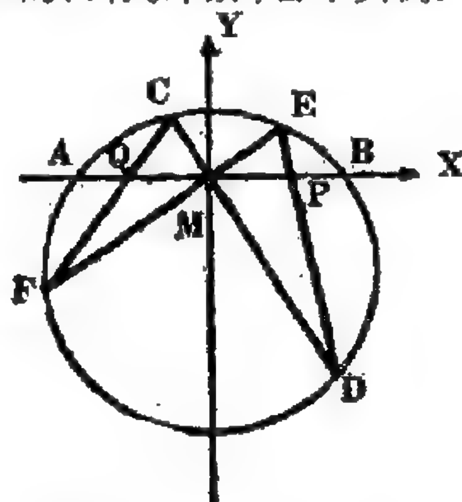


图3-2

要证$PM=MQ$，如果以M为原点，以AB为x轴建立直角坐标系（如图3-2），则只须证明：P与Q的横坐标之和等于0，为此，只要设法寻找一个二次方程，它的一次项系数为零，而点P与Q的横坐标满足该方程。这是可以办到的。

设圆的方程为

$$x^2+(y+a)^2=r^2$$

直线CD和EF的方程依次为

$$y=k_1x, \qquad y=k_2x$$

圆和两相交直线组成的二次曲线系为

$$\mu[x^2+(y+a)^2-r^2]+\lambda[(y-k_1x)(y-k_2x)]=0$$

令$y=0$，则P点和Q点的横坐标满足二次方程

$$(\mu+\lambda k_1 k_2)x^2+\mu(a^2-r^2)=0$$

由于一次项系数为零，故二根x_1与x_2之和为零，所以 $MP=QM$。

这是一个杰出的证明。如果用一个运用综合法的几何证明与之比较，就足以看出这种证法的确是节省了推理步骤。

关于微积分符号，有人提出："莱布尼茨发明的是微积分，还是一个特别巧妙的微积分符号体系？"当然，应当说是：他发明了两者。事实上，他的微分和积分符号抓住了他的微积分的本质，使得符号和概念成为不可分割的。正如人们所说，莱布尼茨的符号表示法可以作为一种"看得见、摸得着的媒介，以便用来指导思维"，最突出的例子是微积分基本公式

$$\int_a^b f(x)dx=F(b)-F(a)$$

（这里$f(x)$是区间〔a，b〕上的连续函数，$F(x)$是$f(x)$的原函数。）

这个公式以极其简洁的符号，揭露了定积分和不定积分这两个不同概念间的内在联系。本来，人们计算定积分，必须计算积分和的极限；对于不同的积分和必须寻求不同的途径，大多是艰难的途径。这种困难，使得微积分方法迟迟不得推广。自从牛顿、莱布尼茨建立这个公式之后，人们就得到了求定积分的一般方法，极大地约简思维，加速了思维进程。正如莱布尼茨自己所说，符号方法可能"对推理增加的威力大大超过光学仪器对视力的辅助作用。"

二、符号的推演代替思维推演

我们知道，莱布尼茨首先提出了仿照数学符号的方式，为

人类的思维也建立一套普遍适用的符号体系，也就是说要创造一个“万能算法”。他认为，数学由于有一套独特的符号系统，因而能够有效地表达思想和进行推理。如果人的思维中的概念、判断和推理等也能用适当的符号来表示，我们就可以借助这套符号来进行思维的运算，用计算来代替思考。“有了这种东西，我们对形而上学和道德问题就几乎能像几何学和数学分析一样进行推论。”“两个哲学家也不需要辩论，因为他们只要拿起石笔，在石板前面坐下来，彼此说一声：我们来算算，也就行了。”尽管莱布尼茨没有完成他的计划，但他的工作对后世有极大的影响。现代，在数理逻辑中，概念、判断、推理和证明全部可以用符号和式子十分简洁地给予表示。例如，古典逻辑中的矛盾律、排中律和传递律可以分别表示如下：

矛盾律：$p\wedge\sim p\equiv 0$；

排中律：$p\vee p\equiv 1$；

传递律：$(p\rightarrow q)\wedge(q\rightarrow r)\rightarrow(p\rightarrow r)$。

在数理逻辑中已经把推理和证明过程符号化、形式化了，从一些简单明了的公理出发，使用一些简单甚至机械的推理规则和步骤，便可解决日常生活或科学研究中的逻辑推理问题。

例 甲、乙、丙、丁四人同时参加一次数学竞赛，赛后，他们四人预测名次的谈话如下：

甲：丙第一名，我第三名；

乙：我第一名，丁第四名；

丙：丁第二名，我第三名；

丁没说话。

最后公布结果时，发现他们的预测都只对了一半，请你说出这次竞赛中，甲、乙、丙、丁四人的名次（假设四人没有并

列的名次)。

解 "甲第一"记为A_1，"甲第二"记为A_2……，以此类推，则甲说的记为C_1A_3，乙说的B_1D_4，丙说的记为D_2C_3。由于他们的预测都只对了一半，因此

$$(\overline{C}_1A_3)+(C_1\overline{A}_3)=1$$

$$(\overline{B}_1D_4)+(B_1\overline{D}_4)=1$$

$$(\overline{D}_2C_3)+(D_2\overline{C}_3)=1$$

故

$$[(\overline{C}_1A_3)+(C_1\overline{A}_3)]\cdot[(\overline{B}_1D_4)+(B_1\overline{D}_4)]\cdot[(\overline{D}_2C_3)+(D_2\overline{C}_3)]$$

$$=1$$

左边展开为

$$\overline{C}_1A_3\overline{B}_1D_4\overline{D}_2C_3+\overline{C}_1A_3\overline{B}_1D_4D_2\overline{C}_3$$
$$+\overline{C}_1A_3B_1\overline{D}_4D_2C_3+\overline{C}_1A_3B_1\overline{D}_4D_2\overline{C}_3$$
$$+C_1\overline{A}_3\overline{B}_1D_4\overline{D}_2C_3+C_1\overline{A}_3\overline{B}_1D_4D_2\overline{C}_3$$
$$+C_1\overline{A}_3B_1\overline{D}_4\overline{D}_2C_3+C_1\overline{A}_3B_1\overline{D}_4D_2\overline{C}_3$$

因为它们的和等于1，故这八个式子中至少有一个为真。

第一个式子$\overline{C}_1A_3\overline{B}_1D_4\overline{D}_2C_3$中，$A_3$和$C_3$即甲和丙都是第三，这个命题不真，故其值为0。第二个式子和第六个式子中，D_2和D_4矛盾，第三个式子中A_3和C_3矛盾，第五个式子和第七个式子中C_1和C_3矛盾，第八个式子中C_1和B_1矛盾，故只有第四个式子$\overline{C}_1A_3B_1\overline{D}_4D_2\overline{C}_3$为真命题，其值为1，所以，乙第一，丁第二，甲第三，丙第四。

三、符号化促进思维的“机械化”

用一个固定的程式解决一类问题，这就是数学机械化的基本思想。追求数学的机械化方法，是我国古代数学的优秀传统之一。我国古代数学研究的中心问题是问题求解，把问题分成若干类，分别研究其解题方法（一系列确定的步骤），会一个方法，便能解决一类问题。《九章算术》就有这种特点。韦达引入代数符号体系之后，代数的中心问题——解方程问题就有了固定的程式，或者说是有了机械法则。对于其它数学问题，特别是几何问题，能不能也建立一套机械化的求解法则呢？

我们知道，笛卡尔首先觉察到符号代数具有一种使思维“机械化”的可能性。他试图建立“万能代数模型”，就是想对一切实际问题建立一套机械化的求解步骤，因为“一切实际问题化为数学问题，一切数学问题化为代数问题，一切代数问题化为代数方程求解问题。”而代数方程求解是有机械法则的。

尽管笛卡尔没有完全实现他的计划，但是，他创立了解析几何方法，把初等几何问题化成了代数问题。

莱布尼茨设想的“万能算法”也是试图实现推理的“机械化”，他曾有过“推理机器”的设想，希望用一台机器代替人的推理活动。莱布尼茨在1671年设计过一台能乘能加的计算机。他说：“像奴隶似的把大量时间花费在计算上，对于聪明的人来说，是很不值得的，因为完全可以把这样的工作委托给其他任何人使用机器去进行。”他竭力寻找《万能算法》就是试图使数值计算乃至人的一切合理的思维过程标准化、“机械化”，从而减轻例行的重复性的脑力劳动。

现代电子计算机的原型当推1936年英国数学家图灵设计的一台计算机，即所谓图灵机。图灵成功的关键就在于对数值计算进行数学抽象和逻辑分析；把思维的推演过程“机械化”。

图灵发现，在数值计算时，计算者是谁，用的什么样的纸和笔等等都是非本质的、无关紧要的。计算过程中实质的东西就是一些符号记在某种器具上，计算的行为随着作为各种结果的各种特定符号而变化。图灵的工作主要是把人们在进行计算时的动作分解为比较简单的动作。我们想象一个人在一张方格纸上做计算，他需要(1) 一种贮存计算结果的贮存器，即纸张；(2) 一种语言，表示加减乘除等操作和数字的符号；(3) 扫瞄区，在计算过程中看到的上下左右几个方格的数字；(4) 计算意向，即在计算的每一阶段打算下一步做什么，例如看到 6 + 9 就要准备进位等；(5) 执行下一步的计算。至于每一步计算，无非是(1) 改变数字或符号；(2) 扫瞄区的改变，往左进位或往右添位等；(3) 计算意向改变等。

另外，图灵还把计算过程转化到一条线性的纸带上进行。例如，26×32的竖式演算，可以改写成26×32=52+780=832。如果每个数字都用二进位数码表示，+－×÷=等运算符号也用二进位数码表示，那么一个计算无非就得到一条纸带上的一串 0 和 1 组成的数串，计算过程就可看作是相继选取符号“0”或“1”的机械动作。

图灵正是根据以上原理设计出图灵机。图灵机可以照人们指定的算法程序进行计算，从而使计算活动机械化、自动化。

以希尔伯特为首的形式主义学派设想把数学化为关于有限符号排列的操作。他们主张使用符号推演代替语言，而符号的使用方法要靠约定的规则。

按照他们的设想，首先要有一个符号表，其中只有有限个符号。这些符号是准备用来代替我们日常讨论数学问题的语言的。

就像用字组成句子一样，符号可以组成所谓的公式。有些公式可能没什么意义，有些公式可能有意义。有意义的公式叫"合式公式"。当然，"合式公式"是用具体的规则来说明的。

"合式公式"相当于命题。从"合式公式"中选择一些基本的公式，相当于公理。为了从"公理"推演出别的"公式"（定理），应当有一些推理规则。这些规则也是明确规定的：即什么符号串可以换成什么样的符号串。

为此，数学推理就变成完全确定了规则的机械化的符号操作了。

对这套符号系统的研究，叫元数学或证明论。

希尔伯特建立了元数学——形式系统的数学之后，进一步考虑了两个问题：

(1) 既然数学命题可以用形式系统的"合式公式"来表示。那么，应用形式的推演规则，能不能推出所有的真命题呢？如果能推出所有的真命题，就说这个系统是完全的。

(2) 形式系统会不会推出矛盾呢？会不会既能推出某一命题，又能推出它的否命题呢？如果推不出矛盾，就说这个形式系统是协调的。

希尔伯特试图证明形式数学系统的完全性与协调性。

尽管希尔伯特的这两个目标都不可能达到。但将数学形式化的基本思想已被大家广泛接受。

形式化的语言可以直接用于计算机程序设计，它的每一步

骤都具有纯粹机械操作的性质。用形式化的语言写的教科书实际上是一串长长的符号链。它经过数学家或机器处理时，就变成另一个符号链。

最近十几年来，已经有人将符号逻辑的研究成果应用于电子计算机，初步实现了数学证明的机械化。我国数学家吴文俊在1977年率先发表了他的初等几何机器证明的方法。

吴氏方法的基本要点是：先把几何问题化为代数问题，再把代数问题化为代数恒等式的检验问题。代数恒等式的检验是机械的，问题的转化过程也是机械的，整个问题也就机械化了。

§4 刺激联想活动

联想是由一事物想到另一事物的心理过程。它是以观察为基础，根据研究对象的特点，结合已有的知识和经验进行想象的思维方法。联想在科学认识活动中起着桥梁作用，它使人们由此及彼、由表及里沟通知识间的联系。牛顿由苹果落地联想到行星运动；阿基米德由浴盆洗澡联想到王冠之谜。在解决问题时，联想得快表现为“点子多”、“反应快”、“头脑灵活”。

符号作为一种可感的实体，是一种刺激，它能激发人们的联想的火花。一个符号形式不仅能刺激大脑，联想到相应符号内容，还能联想到更丰富的情景。毒药瓶上的骷髅头符号能使人联想起中毒、救命车、医院、太平间、火葬场……电视荧屏上暴风雨标志使人联想起大雨、大水、大灾、抗洪救灾……数学符号也不例外，由于数学符号的高度抽象性，它激发的联想更是多方位的、多层次的，高度发散的。例如，符号“1”可算是笔划最少、最简单的一个符号，然而，它却能拨动心弦，

令人产生多方面的联想，例如

$$
1 \longrightarrow \begin{cases} \text{单位圆半径} \\ \sin^2\alpha+\cos^2\alpha=1 \\ \lg 10=1 \\ \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1 \\ \lim\limits_{x\to 0}\dfrac{\sin x}{x}=1 \\ C_n^n=1 \\ \left(-\dfrac{1}{2}\pm\dfrac{\sqrt{3}}{2}i\right)^3=1 \end{cases}
$$

$$
1 \longrightarrow \begin{cases} \text{初始元} \\ \text{单位元} \\ \text{单位矩阵} \\ \text{恒等变换} \\ \text{全集合} \\ \text{接通} \\ \text{真命题} \\ \text{哥德巴赫的“}1+1\text{”} \\ \cdots\cdots \end{cases}
$$

恩格斯曾经对“1”这一单位数量结合初等数学的内容作过很好的阐述。其实“1”在高等数学的各种理论中都有着特殊的涵义，起着重要的作用，现代数学家都很重视“1”这个概念。有的数学家为了给“1”下定义，甚至写了200页纸作准备，可见“1”具有多么丰富的涵义。有些学者从“1”可以联想到几十种、几百种数学表达式。

由于数学符号的刺激，可以联想到相关的定义、公式、基本解题方法及邻近学科的知识。当然，联想的深度与广度，均与大脑这个储存器中的信息量有关，与思维品质的广阔性、深

刻性及灵活性相联系。

按照思维运演的方向来看，由数学符号激发的联想大致可分为纵向、横向和逆向三种。

一、纵向联想

数学符号能刺激大脑，由眼前的数学对象沿纵深方向联想到相关的对象。比如，由数字符号联想到字母符号；由代表数字的字母符号联想到代表抽象元素的字母符号；由几个字母符号联想到多个字母符号；由字母符号的简单组合联想到复杂组合……

不管在理论研究，还是在一般问题解决过程中，这种纵向联想都是不可缺少的。它引导人们逐步深化结论，发展成果。

例如，如果a，b都是正数且$a+b=1$，则

$$\left(a+\frac{1}{a}\right)\left(b+\frac{1}{b}\right)\geqslant\frac{25}{4}=\left(2+\frac{1}{2}\right)^2$$

这个不等式是不难证明的。这里有两个变元，很自然地会联想到三个变元的情况，对于三个正数a、b、c而言，如果$a+b+c=1$，是否会有不等式

$$\left(a+\frac{1}{a}\right)\left(b+\frac{1}{b}\right)\left(c+\frac{1}{c}\right)\geqslant\left(3+\frac{1}{3}\right)^3$$

成立？

可以证明，答案是肯定的。于是进而联想到四个变元、五个变元乃至n个变元的情况。事情果真与想象一样。若$a_i>0$ $(i=1, 2, \cdots, n)$，$\sum_{i=1}^{n}a_i=1$，则不等式

$$\prod_{i=1}^{n}\left(a_i+\frac{1}{a_i}\right)\geqslant\left(n+\frac{1}{n}\right)^n$$

的确成立。

这种深化、推广工作，似乎微不足道，不值一提。其实，数学中不少重大发现都是发端于这种“微不足道”的联想。

举世瞩目的哥德巴赫猜想是怎样引起的？观察下面几个等式：

$$3+7=10$$

$$3+17=20$$

$$13+17=30$$

可以发现它们之间有类似之处：等式的左边3，7，13，17等都是奇素数；等式的右边10，20，30都是偶数；这几个偶数均可表为两个奇素数的和。哥德巴赫正是从“几个偶数”联想到“一切偶数”，根据归纳法推测：“任何一个大于4的偶数都是两个奇素数之和”。也许，哥德巴赫自己也未意识到，如此简单的联想竟会发展为“数学皇冠上的宝珠”。

费马大定理也不是源出于多么复杂的构思。

我国早在商高时代（约公元前1100年）就已经知道不定方程

$$x^2+y^2=z^2$$

至少有一组整数解：$x=3$，$y=4$，$z=5$。

古希腊数学家丢番图求得一般的解答：$x=2mn$，$y=m^2-n^2$，$z=m^2+n^2$，其中m，n（$m>n$）是任意正整数。

如果指数$n>2$，$x^n+y^n=z^n$有没有正整数解呢？费马认为没有，从而得到下面的命题：

当$n>2$时，$x^n+y^n=z^n$没有正整数解。

这就是著名的费马大定理。

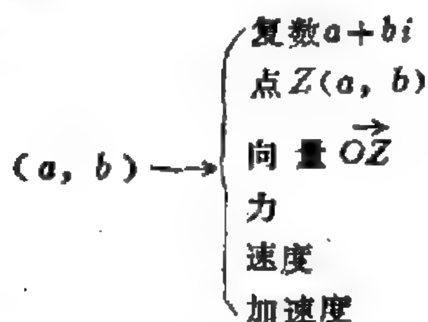
费马曾钻研过丢番图的著作。他在丢番图著作的某页空白处写道：“将一个立方数分为两个立方数，一个四次幂分为两

个四次幂，或者一般地将一个高于二次的幂分为两个同次的幂，这是不可能的。关于此问题，我确信已发现一种美妙的证法，可惜这里空白的地方太小，写不下。”费马正是由“二次幂”联想到一般的“n次幂”。

费马死后，人们才发现了他写的这段文字，可是怎么也找不到这定理的证明，于是激起许多数学家的兴趣，企图给出证明，但都没有成功。布鲁塞尔和巴黎科学院曾设奖金悬赏数次，也得不到结果。三个世纪过去了，费马大定理依然没有得到完全的证明或被推翻。

二、横向联想

数学符号能刺激大脑横向联想。即由眼前的数学对象联想到邻近学科的相关对象。例如有序实数对符号（a，b）可以使人产生如下联想：



这种横向联想，在过去、现在和将来的数学研究中都起着极大的作用。正如H·弗赖登塔尔所说：“数学在其发展过程中，不仅向着其前沿之外扩张，其各个分支还穿过它们之间的边界扩张。……几何问题用代数来解，而代数问题用几何来解。从计算机的机理出发，再一小步就是宇宙的机理或藏在我们头颅之后的思维的机理。”

"虚数"是怎样变得"不虚的"？得益于横向联想。

自1545年，意大利数学家卡当第一个认真地讨论虚数之后，在相当长时期内，人们对虚数抱怀疑态度。卡当自己也觉得奇怪，负数怎么能开平方？他称负数的平方根为"诡辩量"。过了将近一百年，笛卡尔在《几何学》中第一次给这种"诡辩量"取了一个名字叫"虚数"（和"实数"相对）。他认为这种根不是实在的，而是虚的。牛顿也并不认为虚数根是有意义的，这很可能是由于它们"缺乏物理意义"。莱布尼茨虽在形式运算中使用虚数，但并不理解虚数的性质，他把虚数看作"两栖怪物"。直到18世纪，欧拉还是说这种数只存在于"幻想之中"，他在论文《微分公式》中首次使用i表示$\sqrt{-1}$，但很少有人注意它。1801年，高斯系统地使用这个符号，以后才渐渐通行于全世界。

"虚数不虚"的思想是在引进复平面之后才为人们接受的。17世纪时，英国数学家瓦里士已经意识到在直线上不能找到虚数的几何表示。1797年，挪威的测量学家维塞尔向丹麦科学院递交论文《方向的解析表示，特别应用于平面与球面多边形的测定》，首先提出把复数$a+bi$用坐标平面上的点(a, b)来表示，使全体复数与平面上的点建立了一一对应关系，形成了复平面概念。但当时没有受到人们的重视。1806年，日内瓦的阿工在巴黎发表的论文《虚量，它的几何解释》，也谈到了复数的几何表示法。他用"模"这个名词来表示向量$a+bi$的长度，模这术语就源出于此。

高斯在1799年已经知道复数的几何表示，在1799年、1815年、1816年他对代数基本定理作出的三个证明中，都假定了复数和直角坐标平面上的点一一对应。但直到1831年他才对复平

面作出详细的说明，使人们接受了复平面的思想。此后，有些人便把复平面称为高斯平面。

利用复数的几何表示法，复数又可以用坐标平面上的向量来表示。两个复数相加可以按照向量加法的平行四边形法则来进行，一个复数乘以i（或$-i$）相当于表示此复数的向量逆(或顺）时针旋转90°。这就使得物理中的许多向量：力、速度、加速度等等，都可以借助于复数来进行计算，使复数成为物理学和其它自然科学的重要工具。

所谓"笛卡尔连接法"就是源出于代数与几何之间的横向联系。大家熟知，笛卡尔把符号代数应用于几何研究，建立解析几何，这是17世纪的辉煌发现，是数学史上三大发明之一。

狭义地讲，"笛卡尔连接"就是指几何图形与代数方程之间的结合。例如我们一看到$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$，脑子里就联想到椭圆，如果$a=b$，这就是圆。

广义地讲，"笛卡尔连接"是指一切分析的、逻辑的、抽象概念与综合的、直观的、具体形象之间的结合。

看到的	联想到的
$S=ab\ (a,\ b>0)$	a [S] b
$ax+by+c=0$	直线
$\{a_n\}(n=1,\ 2,\ 3,\ \cdots)$	动点a_n
$\lim\limits_{x\to x_0}f(x)=f(x_0)$	曲线在x_0处连续
$f(x)=0$有两个根	曲线与x轴有两个交点
$f'(x)>0\ (<0)$	曲线上升（下降）
$f'(x)=0$	切线平行于x轴

如果说，在解析几何时代需要笛卡尔那样的连接，那么，在科学发展的当代，就更不能忽视这种方法了。科学的发展要求人们在更为广阔的意义上运用这种连接法，更充分地发挥抽象思维和形象思维的互补功能。

1980年杨振宁博士曾在上海作过“爱因斯坦和20世纪下半期的物理学”的演讲，演讲的主题之一就是物理原理几何化。他认为，麦克斯韦用数学方程式表示了法拉第关于磁力线的几何想法，而爱因斯坦在许多文章中都讲过了物理原理几何化问题。爱因斯坦把电磁场看作空时结构，实际上是把它看成几何结构。杨振宁通过非常美妙的想法把引力看成是几何。引力是几何，那么，所有物理原理都可能是几何。应当承认，物理原理几何化从广义上看来也是一种笛卡尔连接。

数学家谷超豪也在一次科学史讲演中谈到了几何学在物理学发展中的作用。他回溯了古希腊几何学家阿波罗内斯的圆锥曲线理论对开普勒发现行星运动三大定律、牛顿发现万有引力定律理论的影响，黎曼的弯曲空间几何理论对爱因斯坦建立广义相对论所起的重要作用，表明了“笛卡尔连接法”的现实合理性。

三、逆向联想

数学符号能刺激大脑逆向联想。

数学中，有些概念是成对出现的。例如，相等与不等、直线与曲线、有限与无限、运动与固定、异与同、收敛与发散等等；有些运算也是成对出现的，例如加法与减法、乘法与除法、乘方与开方、指数与对数、三角函数与反三角函数、映射与逆映射、微分与积分等等。每一对中都含有两个“对立的方面”。

所谓逆向联想，就是由眼前的数学对象联想到对立面一方的对象。例如：

看到的	联想的
−	加法
÷	乘法
“$\sqrt{\ }$”	乘方
log	指数
arcsin	正弦函数
$\int$	微分运算

事实上，作为逆运算的减法、除法、开方、反三角函数、对数函数及积分等总要比原来的运算复杂一些，处理逆运算往往依赖于原来的运算。例如，小学生做除法，是要背着乘法口诀去“凑的”，一见到“÷”号就联想到“乘法”。微积分运算也是如此，求导数只要机械地反复运用导数的四则运算以及复合函数求导法则，而积分运算是以微分运算的逆运算来定义，并没有指出完成积分运算的步骤过程，就种“非构造性”的特征，使得求原函数远比求导数困难。求原函数必须倒回去套用求导公式，这就是说，看到“$\int$”号必然会联想到微分运算。

这种逆向联想在数学研究中具有明显的特殊作用，它是辩证思维方法的一个重要方面。我国古代数学家刘徽用割圆术推求圆周长时，就是“化曲为直”，用直线形来逼近圆。莱布尼茨为了解决在曲线上作切线的问题而引出导数概念，也是用“化曲为直”，在小范围内用直线来代替曲线。牛顿为了解决非匀速直线运动的瞬时速度问题而引出导数概念，他的方法是“化变为不变”，在短时间内用匀速运动来代替非匀速运动。

无论在高等数学中，还是在初等数学中，都充满着各种矛盾，充满着各种对立面的相互转化。恩格斯在数学札记中不但指明了这一点，而且具体分析了一些矛盾，像一和多，直线和曲线等，他不但分析了数学概念之间的辩证关系，而且特别注意到数学运算之间相互转化的重要性。在“数学问题”这则札记中，他以四则运算和代数运算中经常出现的相互转化，如加法和减法、乘法和加法、除法和减法、乘法和除法以及根和幂之间的相互转化等等为例，说明“计算方法的一切固定差别都消失了，一切都可以用相反的形式表示出来。”并且进一步指出：“这种从一个形式到另一个相反的形式的转变，并不是一种无聊的游戏，它是数学的最有力的杠杆之一。”

§5 诱发数学灵感

语言本身就是科学发现的工具，数学语言是数学发现的工具。虚数的发现是很典型的例子。虚数最初是在解二次方程的过程中出现的。1484年，法国人舒开在《算术三篇》中，解二次方程

$$4+x^2=3x$$

得根 $x=\frac{3}{2}\pm\sqrt{2\frac{1}{4}-4}$，他声明这根是不可能的。

1545年，卡当在《大术》中解这样的问题：两数的和是10，积是40，求这两数。

用现代的符号，可列成方程

$$x(10-x)=40$$

$$x^2-10x+40=0$$

得到两个根 $5+\sqrt{-15}$，$5-\sqrt{-15}$。

卡当还发现，某些三次方程的实根竟可以用含负数平方根的代数式表示出来，例如，方程

$$x^3=15x+4$$

的三个实数根 $x_1=4$、$x_{2,3}=-2\pm\sqrt{3}$就可以统一表示成

$$x=\sqrt[3]{2+\sqrt{-121}}+\sqrt[3]{2-\sqrt{-121}}$$

尽管他对负数开平方表示奇怪。但他对这种奇怪的“诡辩量”进行了认真的讨论，并给出了运算的方法。他说：“不管我的良心受到多大的责备，但是，的的确确 $5+\sqrt{-15}$乘 $5-\sqrt{-15}$刚好等于40！”

人们公认虚数是卡当引进的。虚数的引进“照亮了整个世界”，给数学带来了根本性的变革。

怎样创造？怎样发现？爱因斯坦说：“直觉和灵感是头等重要的”。数学符号能暗示信息、刺激联想活动，有助于诱发灵感。数学中很多重大发现都得益于灵感。

一、灵感及其特性

当代，灵感之谜，被誉为“创造学、思维学、心理学皇冠上的一颗明珠”。千百年来，人们对灵感的理论解释众说纷纭，“有多少学者就有多少主张和主义”。古希腊的伟大哲学家亚里士多德说：“灵感就是在微不足道的时间里，通过猜测而抓住事物本质的联系”。G·波里亚把灵感通俗地解释为“好念头”。他写道：

“向求解的突然进展称为‘好念头’、‘妙主意’、‘巧想法’、‘灵机一动’。什么是好念头？是我们观点上的一次重大突变，我们看问题方式的一个骤然变动。在解题步骤方面的一个刚刚露头的有信心的预感。”

G·波里亚把"好念头"作为"灵感"的同义词，他说："想出一个好念头是一种'灵感活动'"，"好念头的出现，每个人都体验过，但只能心领神会而难于言传"。

灵感与直觉相近。简单地讲，直觉就是直接的觉察，它是人们对客观事物的一种迅速而直接的洞察或领悟。灵感是直觉的特殊状态，往往表现为经过一段沉思之后突然产生新设想，思维过程特别迅速，像闪电一样。因此，直觉有时又称为灵感。

灵感的产生常常出现在思考对象已经不在眼前的时候。阿基米德在浴盆洗澡时悟得浮体定律，牛顿从苹果落地悟得万有引力定律，都是得益于灵感。一般说来，某一事物或人物在人们头脑中迅速留下的第一印象，往往被称之为直觉，但这种思维状态均不属于灵感。

钱学森教授在《关于形象思维问题的一封信》中明确指出，"灵感"在创造性思维中是"不同于形象思维和抽象思维的思维形式"。他第一次鲜明地把灵感现象作为人类的一种基本思维形式提了出来。他提倡创立一门"灵感学"。

从思维方式看，灵感思维是从整体上对客观事物作出迅速而直接的判断，它不同于一般的逻辑思维，具有"非逻辑性"。从认识发生看，灵感是一种突发性的创造活动；从认识过程看，灵感是一种突变性的创造活动；从认识成果看，灵感是一种突破性的创造活动。难怪人们经常用"蹦出"、"跳出"、"闯进"、"闪出"和"一闪念"等形象化的名词来描述灵感的"非逻辑性"特征。

二、灵感的触发

灵感的产生带有自发性与随机性，有赖于机会。灵感可以

被外界的偶然的机遇所触发。机遇是“发明家的上帝”。

美国发明家莫尔斯在1832年发明了电报。他创造了至今仍在电报通信中应用的莫尔斯电码，就是用点（短时电流）、线（较长时电流）和空（没有电流）的适当搭配来代表字(字母)和数字。他遇到的最大障碍是远距离传输的时候信号发生衰减现象。他起先采用放大原始信号的方法，但是没有成功。有一天，他搭乘驿车从纽约到巴尔的摩去。他在旅途中观察到，邮车每到一个驿站就要更换拉车的马。突然，他产生了一个想法：在电报线路沿途设置放大站，不断放大信号，终于解决了电报信号长途传输的衰减问题。

美国工程师杜里埃认为，为了保证内燃机有效的工作，必须使汽油和空气能够均匀的混合。可是怎么来实现这种混合呢？这个问题一直纠缠着他。1891年，他看到妻子喷洒香水，于是从这个化妆器具得到启发，创造了发动机的汽化器。其实，汽化器也是一个喷雾器。

珍妮纺纱机的发明过程更为生动，1764年的一天，木工哈格里沃斯与以往一样，又为发明纺纱机的问题伤了一整天的脑筋。傍晚，他疲倦地站了起来，打算暂时丢开这个恼人的问题去做点家务。可是他一不小心，一脚将妻子的纺车给绊倒了。这时，一个现象竟使他看呆了：原来水平放置的纺锤倒过来以后变成垂直竖立了，却依旧在那里转动。哈格里沃斯由此想到，既然纺锤在垂直状态下仍然能转动，那么在纺纱机上并排垂直装上几个纺锤，不就可以一次纺出好几根纱来吗？就这样，他试制成功了新型的“珍妮纺纱机”，大大提高了纺纱效率。

三、数学符号诱发灵感

既然自然标志、自然形象能诱发灵感，作为富有数学思想的数学符号就有更大的诱发力。在数学研究活动中对符号的观察常能诱发灵感。

莱布尼茨研究乘法计算机的原理时，很长一段时间没有找到恰当的解决方法，后来他收到一位法国传教士从中国寄给他的"八卦图"，他从图中符号得到启发，建立了二进制数。

卦的基本符号是爻。爻分为阳爻"——"和阴爻"— —"两种，合称为"两仪"。每次取两个，共有四种不同的排列法，叫做"四象"：

⚌ 太阳　⚍ 少阴　⚎ 少阳　⚏ 太阴

每次取3个，共有八种排列法，叫做"八卦"：

☰ 乾　☱ 兑　☲ 离　☳ 震

☴ 巽　☵ 坎　☶ 艮　☷ 坤

八卦常用来代表八种不同的事物，如东、东南、南、西南、西、西北、北、东北八个方位，或天、地、风、雷、水、火、山、泽八种自然物等。八个方位和八卦对应起来，常画在罗盘的周围。

莱布尼茨将阳爻看作数码1，阴爻看作数码0，于是，"四象"和"八卦"就改写成二进制数：

四象	⚏	⚎	⚍	⚌
二进	0 0	0 1	1 0	1 1
十进	0	1	2	3

如每次取6个爻，可得64种不同的排列，叫做64卦。用上

八卦	☷	☳	☵	☱	☶	☲	☴	☰
二进	0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
十进	0	1	2	3	4	5	6	7

述方法就可把64卦改写成 0 到63这64个二进制数。

如此看来，莱布尼茨创建二进制，正是由于“卦”“爻”符号的诱发。

时过不久，他就研制出了当时最新型的乘法计算机。

也有人提出莱布尼茨发明二进制数是在他知道周易和八卦图之前。莱布尼茨用二进制数原理研究八卦，破译了中国人两千年前创造的这套符号的秘密。这种说法也同样反映了符号诱发灵感的功能——利用二进数符号发现了八卦图的奥密。正如徐利治先生所说，许多数学家都有一种感觉，从符号中得到的东西比输入的更多，它们好像比它们的创造者更聪明。有些符号似乎具备一种神奇的力量，能在其内部传播变革和创造性发展的种子。

面对困难，我们都期望自己的头脑里能闪现出奇妙的灵感。虽然，灵感的产生是不能预期的，但是人们可以通过有意识的思考，去诱发灵感。主要方法是，设法转化问题，多方位、多层次地扩大联想，多渠道地猎取灵感的触发信息，提高机遇概率，一步一步地诱发灵感。猎取触发信息，当然也包括从数学符号中猎取信息。

G·波里亚在《数学的发现》中，提供了一个极为有趣的例子。其中，他细致地描绘了自己的发现过程，试图启发读者领会到灵感是怎样产生的。我们不妨借来考察一番。

已知：如图3-3，三个圆k、m、l具有相同的半径r，且通

过同一点O；l和m相交于A，m和k相交于B，k和l相交于C。

求证：A、B、C三点所确定的圆的半径也是r。

分析　由问题的已知与求证可以看出：圆半径r扮演了一个十分重要的角色，必须在图中引进r。那么，应当在哪儿画出圆半径r呢？由于三个给定圆l、m、k以及三个交点A、B、C的地位是平等的，所以我们没有理由厚此薄彼，似乎应平等对待它们。为此，设圆l、m、k的圆心分别为：L、M、K。连结OL、OM和OK；再把这三个圆心与三个交点分别相连，诸如：LA、KB、MB、KC等。这样就得到一张拥挤的图（图3-4）。

对此，G·波里亚提示说："它有点象老式杂志上的某些画面，这种画，有不止一种效果，如果你按通常的方式去看它，

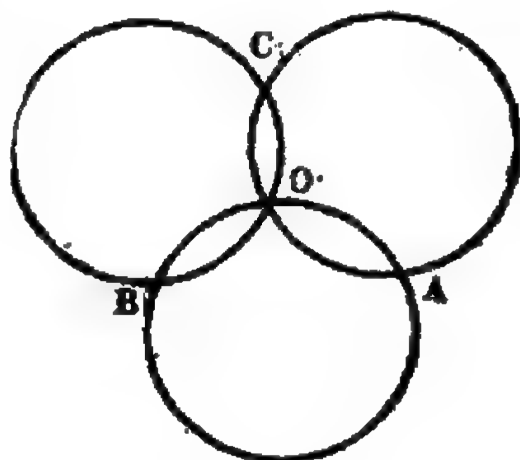


图3-3

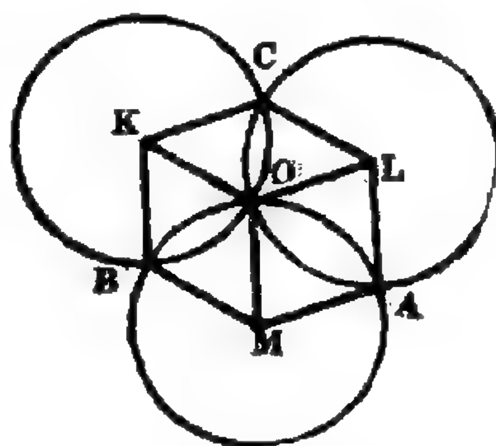


图3-4

它是一个图像，可是如果你转到另一个位置上再换一种特殊方式去看它，那么另一个图像就会忽然闪现在你面前。"他又问："你能从我们这张塞满了直线段和圆的图中看出有第二种含意的图像吗？"

你也许会觉察到隐藏在塞满了的画面里的另一个直线形，

也许会觉察到整个图形实质上是由其中的直线所确定的。因此，我们可以把注意力集中到图3-5上。G・波里亚又说："这个图形是有吸引力的，它使我们想起一些熟悉的东西(想起什么?)"

如果你仍然没有发现什么的话，还可以将图3-5与图3-6作一个比较，这时你又可能想起什么?

显然，图3-6是平行六面体12条棱的一个投影图形，因此通过比较容易想到，图3-5也可看成一个不透明的平行六面体的投影图。这样，原来的问题就解决了。因为，根据立体几何的有关知识可知，如果选择一个平行六面体的投影，使其能看得见的9条棱的投影都等于r（图3-5就是这样的情况），那么，由那个看不见的顶点（如图3-6所示的点E）到A、B、C的三条棱的投影也一定等于r。也就是说，

$$EA=EB=EC=r$$

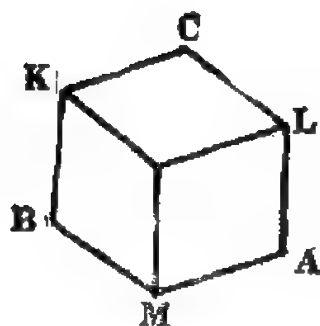


图3-5

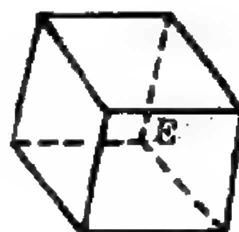


图3-6

这种证法确实非常巧妙，非常独特。一般说来，人们总是把立体几何问题化归为平面几何问题来解决，这里却把平面图形看作是立体的一个投影。正如G・波里亚所指出："如果设想读者遇到了这样的事，从图3-4那些纠缠成一堆的线段和字母中，

平行六面体的形象出乎意料地“跳”了出来，……可能他在某种程度上也将理解灵感这个词的意义，以及怎样去解释好像是一个内在的心声或者一个超自然的神，在启示一样的一个深刻的想法突然出现。”

在这个例子中，G·波里亚通过对几何图形的观察，多方位的观察，获得灵感。几何图形也是符号，正如希尔伯特所说，“几何图形就是直观空间的帮助记忆的符号，所有数学家正是如此来使用它们的。”“算术符号是文字化的图形，而几何图形则是图像化的公式；没有一个数学家能缺少这些图像化的公式，正如在数学演算中他们不能不使用加、脱括号的操作或其它的分析符号一样。”

下面我们再看一个实例。

例　计算定积分$\int_a^b \sqrt{(x-a)(b-x)}\,dx\,(a<b)$。

这个定积分可以用欧拉变换的办法来解决，但计算过程相当繁杂。有没有简便的方法？我们知道，定积分可以表示平面图形的面积。这个定积分能不能与什么平面图形相联系？考虑到被积函数$\sqrt{(x-a)(b-x)}$是$(x-a)$与$(b-x)$的比例中项，读者也许能联想到几何中一条结论，如图3-8，设MN是半圆的直径，$PQ\perp MN$，则

$$PQ=\sqrt{MP\cdot PN}$$

看来，$\sqrt{(x-a)(b-x)}$可以和半圆相联系。这是一个好念头，事实上也是可以做到的。在图3-8中，只要设法使$MP=x-a$，$PN=b-x$，则$PQ=\sqrt{(x-a)(b-x)}$。

因此，得到下面的解法。

如图3-7，设A、B为横轴上的两点，其坐标分别为a、b；

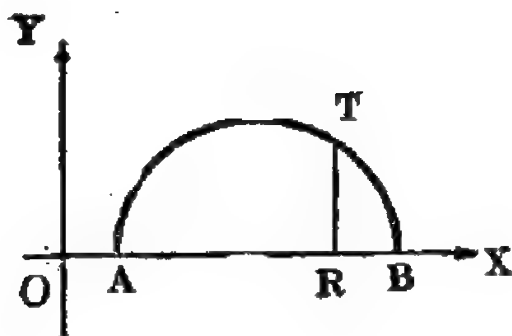


图3-7

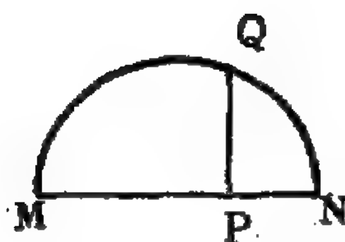


图3-8

R为线段AB上的一个动点，其坐标为x。则$AR=x-a$，$RB=b-x$。以AB为直径作半圆，过点R作$RT\perp AB$交半圆于T，则$RT=\sqrt{(x-a)(b-x)}$。

可见，当动点R从点A移动到点B时，动点T的轨迹就是一个半圆曲线；根据定积分的几何意义，可知所求的积分值就是该半圆的面积，即

$$\int_a^b\sqrt{(x-a)(b-x)}\,dx$$

$$=\frac{\pi}{2}\left(\frac{b-a}{2}\right)^2$$

$$=\frac{\pi}{8}(b-a)^2$$

这种解法是很有特色的，人们往往利用定积分去计算一些平面图形的面积，而这里是利用几何知识，利用平面图形的面积来计算定积分。

一般说来，在数学研究中，那些得益于灵感的方法总是很优美的，而数学符号则是诱发灵感的重要信息源。

第四章　数学符号的创设原则

自1600年以来，数学创造的步幅一直在加大，出现了不少新的数学分支。不过，正如一位学者所说，谈到数学创造，我们自然想到其成果就是有关数学的新体系、新理论、新方法、新概念、新定理、新技术、新公式以及新的应用等等。关于数学符号本身的创造却不易为常人所注意。现在我们已经看到，数学符号与数学发展的关系是如此密切；数学符号的思维功能是如此显赫，研究数学符号的创设原则理应是非常重要，不可忽视的工作。正如L·格莱希尔所讲："初看起来，好像数学范围的扩张必定成为其未来发展的危险之源。……而标记法的改进可能最有效地简化课题和使之易于理解。不但要由数学工作者去探索新的真理，并要由他来设计可以用来发现和表达真理的语言，而在所得到的结果中，一个伟大的数学家也显示出其天才。恰当选择的标记法简化了复杂的理论并把远离的理论联系起来，我对这种标记法所具有的力量有极大的信心，同时，我认为有把握预言，不断增加的关于原则的知识以及不断改进的数学符号的语言，永远能使我们满意地对付仅仅由于这个学科的扩大所带来的困难。"

关于数学符号的创设原则，尚未见到比较系统的论述，这一章，试图在这方面作些探索。

§1 数学符号的形成过程分析

考察历史大有好处。它能帮助我们建立正确的观念，了解客观事物发展的一般规律，展望客观事物发展的趋势和前景。在研究数学符号的创设原则之前，对现有数学符号的形成过程作些历史分析是有益的。

一、数学符号的种类

从数学符号的形成方式来看，现有的数学符号大致可分成象形符号、缩写符号和约定符号三种。

象形符号是用符号的形状特征唤起视觉表象来反映数学概念的符号。数学对象的空间位置、结构或数量关系经抽象概括就得到各种数学图形或图式，图形、图式关系再经缩小或改造就形成象形符号。例如平面图形的符号



等是原型的压缩象形，关系符号

$=$、$\doteq$、$\equiv$、$\backsim$、$\approx$、$\cong$、$>$、$\leqslant$、$\in$、$\notin$、$\supset$、$\vee$、$\wedge$

等是原型的改造符号。

这类符号可由形思义地加以理解、记忆和运用。

缩写符号中的多数是由数学概念的外文词汇的前一个或数个字母构成的缩写，也可以用汉语拼音的类似构造进行缩写。例如

函数 function⟶f

实数集Real number⟶R

极限 limit⟶lim

其它如微分、积分、对数、三角函数、最大最小、和与积的记号

d、$\int$、log、sin、cos、max，min、Σ、π

等均属这一类。

这类符号最初需要以文字概念的记忆为基础唤起听觉表象沟通思维活动，即由音思词、由词及义。

约定符号的形成是与思维活动的习惯和历史有关的，并且具有思维的合理性等特点。例如习惯上用字母x、y、z表示未知数，用a、b、c等表示已知数；用大写字母表示点，用小写字母表示线段或直线，用小写希腊字母表示平面。其它如方幂记号x^n、运算或性质记号+、－等，量词记号$\forall$、$\exists$、阶乘记号！等均属这一类。

这类符号主要应通过规定或约定的简炼性、合理性和习惯性来与思维活动共鸣，由义及形、形义一体地加以理解和运用。

二、数学符号与数学理论的关系

从数学家提出新符号的心理过程来看，数学符号的创设与数学理论和方法密切相关，它总是由于研究某种课题的需要而受到某种数学思想的指引。在这方面最典型例子可算是微积分符号的创设过程。微积分是“一种撼人心灵的智力奋斗的结晶”，M·克莱因在他的《古今数学思想》中曾对莱布尼茨创设微积分及微积分符号的心理过程作过精辟的分析。它对于向往着数学创造的读者来说也许是非读不可的。我们简要介绍如下，从中

可以看到"一个伟大的才智"是怎样"随着他思想的发展而改变他所用的记号"的。

莱布尼茨在《微分学的历史和起源》一文中，把他的微积分思想起源追溯到他的关于数列之和或差的序列的早期工作。

莱布尼茨在1672年到达巴黎后不久，注意到有关数列的下述有趣事实。给定数列

$$a_0,\ a_1,\ a_2,\ \cdots,\ a_n$$

考虑相继两项之差的序列

$$d_1,\ d_2,\ \cdots,\ d_n$$

其中 $d_i=a_i-a_{i-1}$，这时

$$d_1+d_2+\cdots+d_n=(a_1-a_0)+(a_2-a_1)+\cdots+(a_n-a_{n-1})$$
$$=a_n-a_0$$

因此，相继两项之差的和等于原数列的首项与末项之差。

对于平方的序列

0，1，4，9，16，25，36

第一阶差是

1，3，5，7，9，11

第二阶差是

2，2，2，2，2，2

莱布尼茨注意到自然数列的第二阶差消失以及平方序列的第三阶差消失等等另外，他还观察到，如果原来的序列是从零开始的，那么第一阶差的和就是序列的最后一项。

1673年左右，莱布尼茨看到求曲线的切线的正问题和反问题的重要性；他也完全相信，反方法等价于通过求和来求面积和体积。为了把前面所讲的关于序列的一些事实和微积分联系起来，他把序列看作是函数的 y 值，而把任何两项的差看作是

邻近两个 y 值的差。最初他认为x表示序列中项的次序，y 表示这一项的值。

量dx，他经常写作a，这时候等于1，因为它是两个相连的项的序数之差；dy是两个相连项的值的实际的差，然后用omn。(即拉丁文$omnia$的缩写)表示和，并且用l代替dy，莱布尼茨断言$omn.\ l=y$，因为$omn.l$是首项为0的序列的第一阶差的和，这样就给出了最后一项。但是，$omn.yl$产生一个新问题，莱布尼茨获得$omn.yl=\frac{y^2}{2}$的结论是在$y=x$的条件下考虑的。因此，如图4－1所示，三角形ABC的面积是yl的和（对于"很小"的l来说），它也是$\frac{y^2}{2}$。他说："从0起增长的直线，每一个用与它相应的增长的元素相乘，组成三角形。"这几点事实，和一些较复杂的东西，已经出现在1673年的论文中。

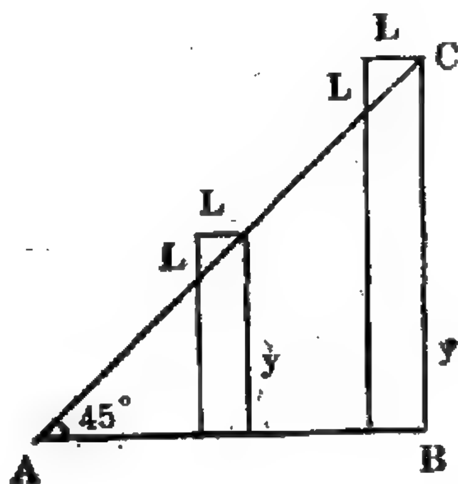


图4-1

下一阶段中，他必须从一串离散的值过渡到dy和dx是x的任意函数y的增量的情况。因为他仍然局限于数列，而在数列中x是项的顺序，所以他的a或dx是1，因此他自由地插入或者去掉a。当他过渡到任意函数的dy和dx时，这个a就不再是1了。但是，当他仍然与和的概念作斗争时，他忽略了这个事实。

因此在1675年10月29日的手稿中，莱布尼茨从

$$omn.\,yl=\overline{omn.\,\overline{omn.\,l}\,\frac{l}{a}} \qquad (1)$$

出发，因为y本身就是$omn\cdot l$，所以(1)是成立的。他用a除l保持量纲，他说，无论l等于什么，(1)是成立的。但是，我们已经在图 4－1 中看到：

$$omn.\,yl=\frac{y^2}{2} \qquad (2)$$

所以，由(1)和(2)得

$$\frac{y^2}{2}=\overline{omn\cdot\overline{omn\cdot l}\,\frac{l}{a}}$$

用近代记号来写，就是

$$\frac{y^2}{2}=\int\left\{\int dy\right\}\frac{dy}{dx}=\int y\frac{dy}{dx}$$

他说，这个结论是值得称赞的。

莱布尼茨从几何的论据中得来的另一个同类型的定理是

$$omn.\,xl=x\,omn.\,l-omn.\,omn.\,l$$

其中l是一个序列中相连两项的差，x是项数，用近代的写法，这个方程就是

$$\int xdy=xy-\int ydx$$

莱布尼茨又假定上式中的l本身是x，可得到

$$omn.\,x^2=x\,omn.\,x-omn.\,omn.\,x$$

他说，但是$omn.\,x$是$\frac{x^2}{2}$（他已经证明了$omn.\,yl=\frac{y^2}{2}$），所以

$$omn.\,x^2=x\frac{x^2}{2}-omn.\,\frac{x^2}{2}$$

由移置最后一项，他得到

$$omn_{\circ}x^2 = \frac{x^3}{3}.$$

在1675年10月29日的手稿中，莱布尼茨决定用∫代替$omn.$；因此

$$\int l = omn.l,\ \int x = \frac{x^2}{2}$$

记号∫是“*Sum*”（和）的第一个字母S的拉长。

可能是由于研究巴鲁的著作关系，莱布尼茨很早就意识到，微分与积分（看作是和）必定是相反的过程；所以面积被微分时，必定给出长度。因此，在10月29日的同一篇手稿中，他说：“已知l及它与x的关系，求$\int l$。”然后，他说：“假定$\int l = ya$，设$l = ya/d$，*……∫意味着和，d意味着差。从已知的y我们总能求出y/d或者l，即y的微差。”

这里，莱布尼茨似乎是在探索∫的运算和d的运算，并看出它们是相反的。他意识到∫实际上是矩形的和，因而是面积的和。所以他承认，要从y回到dy，必须做出y的微差或者取y的微分。然后他说：“但是∫意味着和，d意味着差。”（克莱因评论说：“这可能是后来加进去的”）。因此两个星期以后，为了从y得到dy，他从用d除改成作y的微差，并写作dy。

在此之前，莱布尼茨还认为y值是序列的项的值，x通常作为这些项的序数。但是现在，在这篇论文中他说：“所有这些定理对于那些级数是正确的，在这些级数中项的差和项本身间之比小于任意指定的量。”这就是说，dy/y可以小于任意指定

*）M.克莱因评论说：“这里他把d放在分母中，如果他写成$l = d(ya)$，那就会有更多的意义。

的量。

在1675年11月11日，标题为《切线的反方法的例子》的手稿中，莱布尼茨用∫表示和，x/d表示差。然后他说，x/d是dx，是两个相邻的x值的差，但是显然这里的dx是常数，并且等于1。

根据上面勉强可以理解的议论，莱布尼茨断定一个事实：作为求和的过程的积分是微分的逆。但他并不清楚怎样可以从这样一个粗糙的式子Σydx去得到面积——即怎样从一组矩形得到曲线下的面积。当然，这个困难困扰了17世纪所有的数学家。由于没有清楚的极限概念，或者甚至没有清楚的面积概念，莱布尼茨有时认为后者是非常小而又非常多的矩形的和，因而这个和与曲线下的真正的面积之差是可以忽略的；他有时又认为面积是纵坐标y的和。

关于微分，莱布尼茨在巴斯卡和巴鲁曾用过的特征三角形上建立他的理论。这个三角形（图4-2）由dy、dx和弦PQ组成，他还认为弦PQ是"P和Q之间的曲线，而且是T点的切线的一部分。"虽然说这个三角形是无穷小的，但坚持它和确定的三角形是相似的，即相似于由次切线SU，T点的纵坐标以及切线ST组成的三角形STU。因此，dy和dx是最终的元素，它们的比有确定的意义。事实上，他用三角形PRQ和SUT相似的论据得到$dy/dx=TU/SU$。

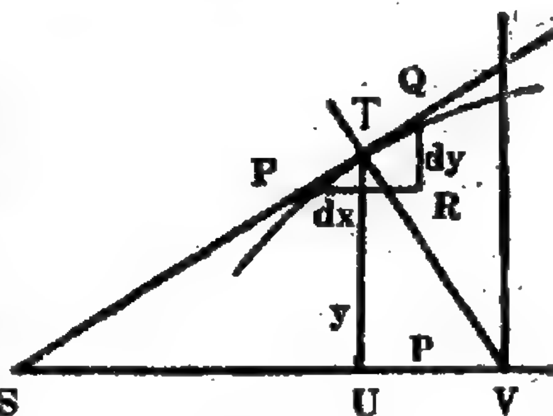


图4-2

在1675年11月11日的手稿中，莱布尼茨说明他怎样能解答

一个确定的问题。他寻求次法线与纵坐标成反比的曲线。在图 4-2 中，法线TV，次法线是UV。由三角形PRQ和TUV的相似性，他得到

$$\frac{dy}{dx}=\frac{p}{y}$$

或者
$$pdx=ydy$$

但是曲线有已知的性质

$$p=\frac{b}{y}$$

其中b是比例常数。因此

$$dx=\frac{y^2}{b}dy$$

所以

$$\int dx=\int\frac{y^2}{b}dy$$

即

$$x=\frac{y^3}{3b}$$

他还解决了其他的反切线问题。

在1676年 6 月26日的手稿中，他意识到求切线的最好方法是求dy/dx，其中dy和dx是差，dy/dx是商。他忽略了$dx\cdot dx$和dx的高次幂。

1676年11月左右，他能够给出一般法则$dx^n=nx^{n-1}dx$和$\int x^n=\frac{x^{n+1}}{n+1}$，其中$n$是整数或分数。他说："这个道理是普遍的，不管$x$的序列是什么样的。"这里$x$仍然意味着序列的项的次序。在这篇手稿中，他说要微分$\sqrt{a+bz+cz^2}$的话，设$a+bz+cz^2$

$=x$，微分$\sqrt{x}$，然后乘以dx/dz，这就是链式法则。

1680年，dx成为横坐标的差，dy成为纵坐标的差。他说："……现在把dx和dy取为无穷小，或者把曲线上的两个点理解为它们中间的距离比任何给定的长度都小……"

他把dy叫做当纵坐标沿着x轴移动时y的"瞬刻的增长"。但是图4－2中的PQ，仍被认为是直线的部分。它是"曲线的一个元素，或者是代替曲线的无穷多角的多边形的一条边。……"他继续用通常的微分形式。例如，如果$y=a^2/x$，则

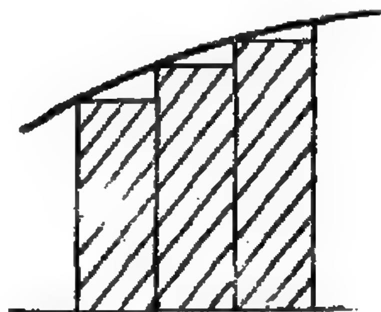


图4－3

$$dy=-\frac{a^2}{x^2}dx$$

他还说差是相反于和的。因此为了得到曲线下的面积(图4-3)，他就计算矩形的和并说能忽略剩余的"三角形，因为它们同矩形相比是无穷小……，因此在我的微分中，我用$\int ydx$表示面积……。"

尽管他先前说过，dx和dy是很小的差，但他仍然谈到序列。他说，"差与和是彼此相反的，这就是说，一个级数〔序列〕的差之和是级数的项，级数和的差也是级数的项，所以我用$\int dx=x$表示前者，用$d\int x=x$表示后者。"事实上，在1684年以后写的手稿中，莱布尼茨说他的无穷小方法已经众所周知地作为差的微积分了。

莱布尼茨在微分方面的首次发表的文章是在1684年的《教师学报》上。

对于dy、dx和dy/dx的最终的含意，莱布尼茨仍然是含糊的。他说dx是两个无限接近的点的x值的差，切线是连接这样两点的直线。虽然他在各阶无穷小量之间确实作出了区别，但是他没有经过证明就仍掉了高价微分。

正如M·克莱因所说："莱布尼茨的工作，虽然是如此的零碎不全，以致几乎不能理解，但它富有启发性而且意义深远。""莱布尼茨煞费苦心地工作，要把记号选得最好。他的dx，dy和dy/dx仍然是标准的。他引进记号$\log x$，对于n阶微分引进d^n，甚至对∫与n重和分别引进d^{-1}与d^{-n}"。他为人类作出了杰出的贡献。

三、数学符号的演变

从数学符号的演变过程来看，数学符号总是在交流传播过程中不断改进的，它要受到文化、经济、技术和印刷条件等等社会背景的影响，有的甚至还经历过戏剧性变化。

例如根号"$\sqrt{\ }$"，18世纪有些学者(如欧拉)猜想"$\sqrt{\ }$"是$radix$(拉丁文的根字)的字头r的变形，后来经过仔细研究，证明不是。原来德国人在1480年前后，用一个点"·"来表示平方根，如·3就是3的平方根，‥表示4次方根，而…表示立方根。到16世纪初小点带上一条尾巴变成✓，像个小蝌蚪。数学史家推测，这可能是写快时带上的。"小蝌蚪"在书写、运算及印刷等各方面毕竟是有缺陷的，1525年路多尔夫的代数书用√8，√√8、√√√8分别表示8的平方根，四次方根及立方根。

我国数学史上零号的演变过程中，也表现出某种偶然性。我国古代用空格表示零。古书缺字都用□来表示，数字间的空

位，自然也可以用□来表示。在书写的时候，字体常写成行书，而方块也就容易划成圆圈了。以0作零号，最早在金《大明历》（1180年）中见到。如403写作“四百〇三”。到秦九韶的《算书九章》（1247年）就大量使用〇号，如3076800记作丨||〇ㄒ丄Ⅲ〇〇，有人以为〇号是从印度传过来的，这是没有根据的。我国零号的演变过程是清楚的，而且起初的写法和阿拉伯数码的扁圆0不同。

*　*　*　*

数学符号的形成过程如此复杂，似乎就没有什么创设原则可谈了。其实不然，数学符号的产生决不是自生自灭的，也不是数学家随心所欲的。我们已经看到，数学符号的创造总不会脱离数学理论和方法去孤立地进行，他要受到数学思想的影响。散见在各种数学著作中的许多片断表明：对于优秀符号的创设，数学家总是自觉地或不自觉地、或多或少地遵循着一些科学原则。正如美国两位数学家所说：“适者生存”，一切数学符号，凡是符合这些科学原则的才得以“生存”，否则，必然被淘汰。足见，我们从中总结出数学符号的创设原则不仅是必要的，也是完全可能的。

§2　创设数学符号的基本原则

数学符号的功能是用于传播数学思想，认识数学世界的奥秘。因此，创设数学符号的总的原则是：有利于数学思想的交流、传播；有利于发挥其思维功能；有利于数学的发展。具体地讲，数学符号的创设，应该遵循下面几个基本原则。

一、确定性原则

我们知道，任何符号都包括两个方面，即能指与所指，符号功能是建立在能指与所指之间的约定关系基础上的。“能指与所指之间的关系，在任何情况之下都是约定的。”包括按符号使用的自然标志。数学符号也不例外，数学符号的意义是需要解释的，解释符号就是使它同某些具有现实意义的概念或心智想象相联系。所谓象形符号和缩写符号也是约定俗成的符号。

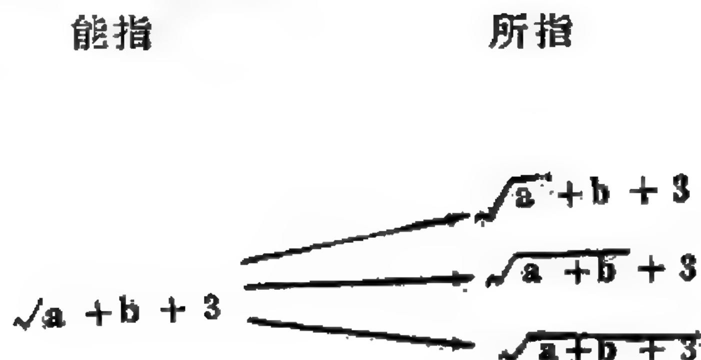
约定可以明显的，也可以不明显的。另外，约定概念是相对的，约定有程度上的区别：它可以是很严格的，也可以是不怎么严格的；可以是一致的，也可以是不大一致的；可以是受某种限制的，也可以是不怎么受限制的。在数学符号中，能指与所指之间的关系的约定是严格的、几乎是一致的。

数学作为一门要求十分精确的学科，它以逻辑的严格性和结论的确定性而著称，要求一种准确性很强的语言与之相适应。“数学中不存在那种意义含混的词”。因此，数学要求每一个符号都有确定的含义。

在自然语言中，一个能指可以指向好几个所指。例如，“号”作为能指，其所指可以是乐队用的一种乐器；也可以是一个“大声呼喊”、“大声哭”的动作。但在数学中，“成功的研究依赖于构想合理的标记”，而“合理的标记法，总要不含糊、不混乱地表达基本数学的本质”。因此，在数学符号中每个能指只能指向一个所指。这就是含义确定性原则。

在§1中曾提到，1525年路多尔夫的代数书用✓8，w✓8，w✓8分别表示8的平方根，4次方根与立方根，这种标记法就不符合确定性原则。因为，当被开方数有好几项时，一个能指

就可指向几个所指。例如：



如此，失去了含义确定性。为了避免混淆，笛卡尔在有关的几项上面用括线把它们连起来，才成为现在的根号形式。

在牛顿所创设的微积分符号中，也有不少含混的地方，最突出的例子是，他用 0 表示瞬（无限小增量）。这个标记与通常的零号容易混淆，“0”究竟是不是“零”？牛顿认为不是；既然这样，为什么在运算中可以略去含有 0 的幂的项呢？牛顿没有给出合乎逻辑的说明。

必须指出，所谓“每个能指只能指向一个所指”是相对于同一系统而言。例如，符号“1”和“0”在集合代数中均有确定的含义——“全集”和“空集”；如果在命题代数中，它们则分别有另外的含义——“真命题”和“假命题”。因此，同一数学符号可以作多种意义的解释，只要这些解释能具体说明数学符号表示的抽象物与其具体的关系。这与数学符号含义的确定性是不矛盾的。

二、简明性原则

数学语言的重要特征之一就是简洁性。在数学中，叙述一个

概念在准确、完备的前提下，力求简洁，话虽不多，但意思能表达得清楚。使用数学语言常常可以大大缩短自然语言表达的“长度”，使之简洁清楚。为了保证数学语言的简洁性，数学符号的创设就应该遵循表达洁明性原则。

简明性原则是由符号的本质所决定的。符号是传播意愿的标志，既然是标志，就力求表达简明。“简单明了”是美的体现。皮亚诺曾指出：“在两种符号体系中，符号用得较少的一般是更可取的。”数学史表明，一切数学符号的演变过程总是服从着逐步简明化的规律。

罗马数码在表达方面不具有洁明性，“一个简单的数要写成长长的一串”。因此，印度-阿拉伯数码问世后，罗马数码必然要被淘汰。任何一种新的改革，总会遇到阻力。印度-阿拉伯数码也曾遭到一些人的猛烈反对。其理由是：这些数字“奇形怪状”，长时期未能标准化，零这个数字尤其容易弄错；要耐心学会很多道理后才能正确使用这些新的符号。另外，当时的反对者还有一个“更为有力、颠扑不破”的理由：新数字太容易弄虚作假了。0很容易改为6或9，1也容易改成4，6，7或9，其他数字的形式也可以窜改，往往还可以在已经写下的几个数字中间或后面插进一些数字。正是由于可能发生这些弄虚作假的现象，1299年佛罗伦萨的法令中就禁止在金融事务中使用印度-阿拉伯数码而强迫使用罗马数码。

然而，印度-阿拉伯数码的种种优点，给人以挡不住的美的感受，它最终还是为人们所普遍采用。

用算筹表示数目，在我国数学史上曾起过很大的积极作用，但是，它与印度-阿拉伯数码相比，在表达方面就显得不够简明。

关于分数符号，我们在第一章§3已看到，在阿尔·哈萨的著作中，用

$$\frac{3}{5}\ \frac{3}{8}\ \frac{2}{9} \qquad \text{表示} \frac{2+\dfrac{3+\dfrac{3}{5}}{8}}{9}$$

这种表达是不够简明的，说准确一些是“简而不明”。

在斐波那契《算盘书》中，式子从右到左，整数部分写在分数的右边，将$12\frac{1}{2}x$ 写成 “radices$\frac{1}{2}$12”，14世纪中叶还有用 $3\,\bar{5}$ 表示$\frac{3}{5}$的。和现行的分数线符号相比，可以看出这些表达都是不够简明的。

分指数最早在奥力森的《比例算法》中出现。他把

$$2^{\frac{1}{2}}\ \text{写作}\boxed{\frac{1\,.P}{2\,.\,2}}$$

$$\left(2\frac{1}{2}\right)^{\frac{1}{4}}\text{写作}\boxed{\frac{1\,.P\,.\,1}{4\,.\,2\,.\,2}}$$

有时，他也把

$$9^{\frac{1}{3}}\ \text{写作}\frac{1}{3}\,.\,9',\ 2^{\frac{1}{2}}\text{写作}\frac{1}{2}.2'.$$

这些写法，因为表达不简明，注定要被淘汰。

J·模棣尔曾批评过那种繁琐的标记法。他认为从传播的角度来讲，不应该在理解的方式上给对方加重负担。

数学符号用以表示数学抽象物。数学抽象物具有抽象层次上的区别。例如，不同基数的集合、不同价的微分、不同维度

的空间之间的运算关系等等。因此，合宜的数学符号系统，不仅要能反映同一层次上不同数学抽象物的差别，还要能反映不同抽象层次的差别，即显示出不同的抽象度。因此，表达的简明性，也包括能简明地表达各层次的数学抽象物。在微积分时代，牛顿及许多英国数学家使用$\dot{x}$，$\dot{y}$，而莱布尼茨及许多德国数学家使用dx，dy，“点主义”与“d主义”两派争论了好些年。大家知道，最后，莱布尼茨的符号占了上风。其原因之一是因为他的符号体系更适于表达高阶导数和高价微分，而且可以由正整数阶推广到负数阶和分数阶，由此导致了运算微积的发展，对后来的数学研究产生了很大的影响。

三、方便性原则

方便性原则是指数学符号的创设要考虑到使用的方便，它包括读、写、运算及推理的方便，甚至还包括便于利用计算器械。

方便性原则也是由符号的本质所决定的。数学符号作为交流数学思想工具，其重要价值就在于使用的方便性。

罗马数字对于加、乘运算就极不方便；用算筹表示数字，其运用的方便性也不及印度-阿拉伯数字。

数e为什么“可贵”，其原因之一就是自然对数$\ln x$的求导运算方便，由此又为高等数学乃至一切自然科学带来一系列的方便。

行列式与矩阵的引入，其目的是为了“速记”，也就是为了方便。

二进制为什么得宠，其原因就是它便于电子计算机采用。

首先，二进制只需 0 和 1 两个数码就可以表示一切数目。

这对于机器来说最为有利，可以大大简化计算机的“运算器”。每一数位上的数字1和0最适宜表示为电路中电气信号的存在或消失。例如，“1”表示为电压脉冲的存在，“0”表示脉冲的消失。这样，二进位制的任何数就可以用排成一定顺序的脉冲表示。因此，只要找到一种具有两种稳定状态的元件就可以用二进制表示数了。而这种元件是很多的，例如：电灯的“开”与“关”，磁蕊的“充磁”与“消磁”，纸带的“穿孔”与“无孔”，晶体管的“通导”与“截止”等等。如果在计算机上采用其他进位制，则需要具有多种稳定状态的元件，比如十进制就需要具有十种稳定状态的元件，这在技术上是很困难的，而且线路复杂，造价昂贵。

其次，二进制的运算比较简单，可以大大提高运算速度，这对于电子计算机来说是至关重要的。从二进制的加法表与乘法表

+	0	1
0	0	1
1	1	10

·	0	1
0	0	0
1	0	1

可以看出：对每一位而言，二进制的加、乘运算都是仅有四种情形，但是十进制的加、乘运算却各有100种情形，由此可见运算的繁简。

另外，二进制不但可以进行数值运算，而且还可用以进行二值逻辑运算。把“1”看作是真命题的逻辑值，“0”看作是假命题的逻辑值；规定逻辑加、逻辑乘与逻辑非的运算法则为：

逻辑加

⊕	0	1
0	0	1
1	1	1

逻辑乘

⊙	0	1
0	0	0
1	0	1

逻辑非

A	$\overline{A}$
1	0
0	1

这样，电子计算机就能进行判断和推理，机器也就有了人类思维的某些功能。

二进制的“1”与“0”还可用以表示生物神经系统的兴奋与抑制两种状态，为此，神经系统就可与电子线路中的二进制网络类比。以离散的二进制元件为基础的电子计算机，就成了神经系统某些功能机制的理想模型。

一种数学符号是不是便于使用，决定了它的生命力。数学符号和一切劳动工具一样，总是在使用方便性上逐步改进。这已成为诱导数学符号演变的无形的指挥棒。我们知道，我国古代，数字间的空位用□来表示，在书写的时候，字体常写成行书，方块划成了圆圈○。乍看起来，0号的形成似乎带有偶然性。其实，偶然性后面常常隐藏着必然性，人们为了写得快，而把□写成圆圈○，这本身就说明在使用的方便性上圆圈胜过方框。虽然是无意的，但圆圈一经出现，必然取代方框。“无意插柳柳成荫”，这与柳的生命力有关。

四、启发性原则

为了能运用数学符号“有效地处理新的事物，把握其意义和价值”，认识数学世界的奥秘，要求数学符号富有暗示作用，有利于刺激联想、启发思维。这就是创设数学符号的启发性原则。

象形符号，如$\Box$、$/\!/$、$\odot$等等，都是富有启发性的，人们看到这些符号形式，就想到符号内容——它们所代表的数学抽象物。正如希尔伯特所说："新符号必须服从于新概念。我们用这样的方式来选择这些符号，使得它们会令人想到曾经是形成新概念的缘由的那种现象。"

关于两个三角形相似的符号

$$\triangle ABC \backsim \triangle EFG$$

在现代的书刊中总是默认：两三角形顶点按照字母符号的书写次序相对应，即A对应于E，B对应于F，C对应于G。但在古代书籍中，并没有引入这种关于次序的附加规定；为了判断哪个顶点和哪个顶点相对应，读者必须看图或记住其推导情况。

现代记法显然大大优越于古老记法。使用现代记法，富有启发性，我们不必看图就可由公式得出很多结果。例如，我们可推出

$$\angle A = \angle E$$

$$AB:BC = EF:FG$$

以及其它这类关系。

细想一下，现代记法和古老记法的差别，仅仅是次序不同而已，然而，现代记法更充分地反映了事物的次序和联系，它就可能提供丰富的含义。可见，在启发性方面精心设计符号是绝对必要的。

莱布尼茨总是竭力使它的符号表示法富有启发性。他写道：

"为我们提供一条阿里阿德涅线*，即一种看得见、摸得着的媒介，以便用来引导思维，就好像几何学中所画的图形和为

初学算术的人建立的运算公式那样。"

莱布尼茨所创设的微分符号的确达到了他的目标，可以作为一种“看得见、摸得着的媒介，以便用来指导思维”。下面提供几个实例。

使用拉格朗日引入的函数符号，我们可以把求导运算的链式法则表达为：如果$h(x)=f(g(x))$，则

$$h'(x)=f'(g(x))g'(x)$$

不过，该公式中的符号既没有说明这个公式为什么成立，也没有说明怎样来证明它。如果使用莱布尼茨的微分符号，设$z=f(y)$，$y=g(x)$，则上述公式成为

$$\frac{dz}{dx}=\frac{dz}{dy}\cdot\frac{dy}{dx}$$

对比之下，这个公式显得富有启发性：公式本身显然会使我们想到它是成立的，因只须把右端的两个微分dy相消，就好像它们是实数一样。另外，这样消去微分符号的想法还会引导我们想到这个公式的合乎逻辑的证明方法，即用有限增量Δx、Δy、Δz来代替微分dx、dy、dz，然后取极限。现代很多数学分析课本中就是这样证明的。

链式法则的积分形式是代换积分公式

$$\int f(g(x))g'(x)dx=\int f(u)du$$

由符号代换$u=g(x)$，$du=g'(x)dx$可以相信这公式成立，而不论它的证明如何。这就说明微分形式$f(u)du$在任意变量变

* 阿里阿德涅是希腊神话中克里特王弥诺斯的女儿。传说雅典每年必须进贡7对童男童女，作为关在迷宫中的牛首人身怪物弥诺陶罗斯的食物。英雄忒修斯作为贡品来到克里特，阿里阿德涅对他一见钟情，给他一个线团，让他拿着线团进入迷宫，杀死弥诺陶罗斯后循线返回，后来，人们常用“阿里阿德涅线”比喻摆脱困境的办法。

换下的不变性，这是莱布尼茨的重要发现之一。

还有一个典型的事例。现在考虑由曲线$y=f(x)$绕x轴旋转而生成的曲面。这时，莱布尼茨把曲线的无穷小的一段ds看作是两直角边为dx、dy的"特征三角形"的斜边（见图4-4），由勾股定理得到

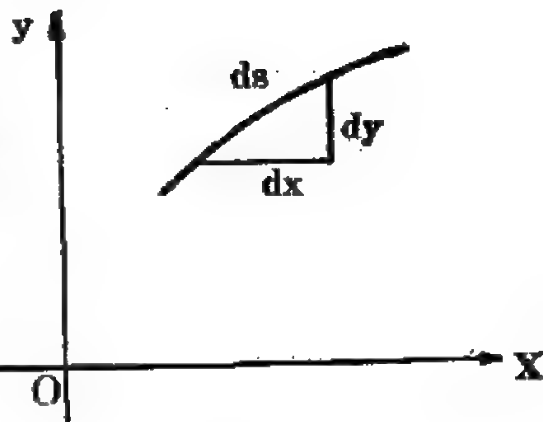


图 4-4

$$ds=\sqrt{(dx)^2+(dy)^2}$$
$$=\sqrt{1+(\frac{dy}{dx})^2}\ dx$$

当ds绕x轴沿半径为y的圆旋转时，就生成无穷小的面积

$$dA=2\pi yds$$
$$=2\pi y\sqrt{1+(\frac{dy}{dx})^2}\ dx$$

把这些无穷小的面积相加，就得到所考虑的曲面的面积

$$A=\int dA=\int 2\pi y\sqrt{1+(\frac{dy}{dx})^2}\ dx$$

可见，使用莱布尼茨的符号，经过程式化的、合乎情理的推理，便可得到求曲面面积的公式。其实，要严格证明这个公式还有些麻烦，它要求对曲面面积的概念进行详细讨论，给出严格的面积的定义，然后证明它和上述公式是一致的。

正因为莱布尼茨的微积分符号富有启发性，因此，"它所释放的能量要比投入的多"。

五、和谐性原则

和谐性是数学美的重要标志之一。和谐的概念最早就是由毕达哥拉斯学派以数学的观点研究音乐而提出的。他们发现音乐在质的方面的差异是由声音在量的方面的比例差异决定的，因此，认为音乐是对立因素的和谐的统一。美就是由杂多导致统一，由不协调导致协调。数学家非常重视他们的方法和理论是否优美，那么到底是什么使我们感到优美呢？彭加勒指出："那就是各个部分之间的和谐、对称和恰到好处的平衡。" 量子力学创始人海森堡也说过："美是一个部分与另一个部分及与整体的固有的和谐。"作为现代数学三大特点之一的严谨性，正是反映了数学体系的和谐美。

数学的语言是由符号组成的，数学符号的创设必然要遵循和谐性原则——考虑各种符号之间的和谐关系。

对于某些有关联的数学抽象物要赋之符号，特别要注意符号之间和谐性。数学中，很多数学概念的意义，是随着数学的发展而逐步变化，逐步丰富的。为了使概念适用于更大的范围，就必须扩展原有的概念，重新给它定义。其间，我们赋于的一系列符号就必须是相互和谐的。

例如，指数符号的创设。韦达对数学符号颇多改良，但没有创设优良的指数符号。西方最早提出负指数的是英国瓦利斯，在他《无穷小算术》中有这样的话：

"平方数倒数的数列$\frac{1}{1}$，$\frac{1}{4}$，$\frac{1}{9}$，…的指数是－2，立方数倒数的数列$\frac{1}{1}$，$\frac{1}{8}$，$\frac{1}{27}$，…的指数是－3，两数逐项相乘，就得到'五次幂倒数'的数列$\frac{1}{1}$，$\frac{1}{32}$，$\frac{1}{243}$，…。它的指数显然

是$-2-3=-5$…同样，'平方根倒数'的数列$\frac{1}{\sqrt{1}}$，$\frac{1}{\sqrt{2}}$，$\frac{1}{\sqrt{3}}$，…的指数是$-\frac{1}{2}$，…"

这是一个很大的进步，不过瓦利斯没有真正使用2^{-2}，2^{-3}，$2^{-\frac{1}{2}}$的指数符号，只是说$\frac{1}{4}$，$\frac{1}{8}$，$\frac{1}{\sqrt{2}}$的指数是-2，-3和$-\frac{1}{2}$。

现行的分指数和负指数符号是牛顿创设的。他在1676年6月13日给信莱布尼茨，里面说到：

"因为代数学家将aa，aaa，$aaaa$等写成a^2，a^3，a^4等等，所以我将$\sqrt{a}$，$\sqrt{a^3}$写成$a^{\frac{1}{2}}$，$a^{\frac{3}{2}}$；又将$\frac{1}{a}$，$\frac{1}{aa}$写成a^{-1}，a^{-2}，…"

这里，瓦里斯与牛顿考虑了相关概念之间以及运算中的和谐关系。

考虑和谐性，不仅是创设数学符号的客观要求，也是创设数学符号的一种成功的经验、方法。

"$0!$"是怎样引出的？当n是自然数时，"$n!$"表示从1开始的n个自然数的乘积，即

$$n!=1\cdot 2\cdot 3\cdots(n-1)\cdot n$$

按这规定，当$n=0$时"$0!$"显然没有意义。我们知道，任何符号总包括两个方面，一是符号形式，二是符号内容。有些计算要求我们重新规定$0!$的意义，也就是要给符号形式$0!$赋于符号内容。怎样规定呢？

考察公式

$$C_m^n = \frac{m!}{n!\ (m-n)!}$$

这里m、n是自然数，且$m>n$，当$m=n$时。

上式左方为$C_n^n=1$，右方出现$\frac{m!}{m!\ 0!}$。

为了使$m=n$时，公式仍旧成立，就必须规定$0!=1$。因此，规定$0!=1$纯粹是考虑和谐性的必然结果。

妙得很，这种规定还保证了更高层次中的和谐性。

18世纪，由于研究插值理论与反微分两个问题，欧拉考察了积分

$$\int_0^\infty x^{a-1}e^{-x}dx \quad (a>0)$$

勒让德称它为伽码函数，用$\Gamma(a)$表示。

即：
$$\Gamma(a)=\int_0^\infty x^{a-1}e^{-x}dx$$

伽码函数是继初等函数之后，在分析及分析应用中最重要的函数之一。

利用分部积分，可得

$$a\int_0^\infty x^{a-1}e^{-x}dx$$

$$=x^a e^{-x}\Big|_0^\infty+\int_0^\infty x^a e^{-x}dx$$

这就是说

$$\Gamma(a+1)=a\Gamma(a)$$

重复应用这个公式，就给出

$$\Gamma(a+n)=(a+n-1)(a+n-2)\cdots(a+1)a\Gamma(a)$$

利用这个公式，无论是对于多么大的a值来计算Γ，总可以化

为是对于$a<1$来计算Γ。

在这式中取$a=1$，并注意到

$$\Gamma(1)=\int_0^{\infty}e^{-x}dx=1$$

那么就发现$\Gamma(n+1)=n!$，也就是得到著名公式

$$n!=\int_0^{\infty}x^n e^{-x}dx(n为自然数) \qquad (1)$$

这项发现的重要性，主要是它给出了量$n!$的分析表达式。很大的数的阶乘，无论在理论研究上，或者在实际的计算上，都起着重要的作用，但是，因为很大的数的阶乘，按照它的定义，是一个很复杂的不便于估计数值的东西，不要说它的精确数值，就连它大小的级人们也无法直接求得，所以无论对于理论或者实际应用来说，当n很大时，能找出一个$n!$的既简单而又便于估计的近似表达式，是一件很重要的事情，这个著名公式为这项工作奠定了基础。

幸运的是，这个公式在$n=0$的情况下也还是成立的。这就是说，规定“$0!=1$”与该公式也是和谐的。这并不是偶然的巧合。不同数学概念之间，存在着内部的联系，按成功的经验进行创造往往是很顺利的。正如希尔伯特所说，数学家们所经常感觉到的那种令人惊讶的相似性和仿佛事先有所安排的协调性，其根源就在于思维与经验之间的相互作用。

§3 为实际需要而创造

我们已看到莱布尼茨创设微积分符号，完全为了研究微积

分方法的需要，其间，他随着自己思想的发展，不断更新他的符号，微积分方法的发明与微积分符号的发明表现出一种同步过程，难怪有人问："莱布尼茨究竟是发明了微积分还是微积分符号?数学史表明，一切符号都不是数学家凭空创造出来的，而是由于数学研究的需要，伴随着数学方法的创新而创造的。如果把这个也作为一条原则的话，那就是要为数学研究的实际需要而创造。

伽罗瓦曾说过："新颖问题需要使用新名称、新符号。我不怀疑，这种不方便在开头的时候将使读者产生反感，他们很难原谅这种生疏的语言，即使作者是他们素所景仰的人。但是归根到底，我们只好适应题目的要求，因为题目的重要性值得注意。"伽罗瓦建立方程式可解性理论的重要关键就在于他对置换群引进了一系列重要概念（如"正规子群"、"单群"、"复解"以及群与群之间的"同构"等）及其相应的符号。他开辟了代数学的崭新领域——群论。在数学史上群论的发明可以与解析几何、微积分的发明比美。伽罗瓦被誉为数学史上的头等天才。人们谈到数学的发明创造总会谈到伽罗瓦理论，我们谈数学符号的创造也不能放过它。

事情必须从代数方程的根式解法谈起。

代数基本定理从理论上肯定了代数方程根的存在性，然而，它的各种证明都不是构造性。没有指出求根的一般途径。那么，怎样求出方程的根呢?

二次方程的解法，早在远古时代已解决，三次、四次方程的解法是在16世纪解决的，这些方程的解法有一个共同的特点，它们的根都可以用它的系数的代数式来表示，所谓代数式，就是只含有限次的加、减、乘、除和开方五种代数运算的表达

式。因此，人们说以上方程可以用代数方法求解，或者说以上方程均有根式解。自从三次、四次方程的根式解法解决后，五次方程的根式求解问题自然便提到议事日程上来。它吸引了众多的著名数学家，包括欧拉在内。但是两百多年来的努力都落空了。他们认为五次方程的求根公式是存在的，只是没有找到而已。

到了18世纪下半叶，拉格朗日精心分析了二次、三次、四次方程的根式解结构之后，引进了方程的预解式概念。

以二次方程为例。大家熟知，二次方程

$$x^2+bx+c=0$$

的求根公式是

$$x_1=\frac{1}{2}(-b+\sqrt{b^2-4c}),\quad x_2=\frac{1}{2}(-b-\sqrt{b^2-4c})$$

拉格朗日发现$\sqrt{b^2-4c}$与两根x_1、x_2有密切联系，它可表成$\sqrt{b^2-4c}=x_1-x_2$。简记$\phi(x_1, x_2)=x_1-x_2$，则ϕ即称为二次方程的预解式。为什么叫预解式？因为根据韦达公式

$$\begin{cases} x_1+x_2=-b \\ x_1\cdot x_2=c \end{cases}$$

只要能将ϕ(即x_1-x_2)用b，c表出，则由$x_1+x_2=-b$立即可得二根x_1，x_2。

拉格朗日看出，预解式和方程诸根的置换性质有关。他拟定了一种解n次代数方程的方案，并用他的方案成功地、统一地处理了二、三、四次方程的根式解问题，可是，当他用这种方法去解五次方程时碰了壁，最后只能被迫得出结论说，一般高次方程($n>4$)的根式解看来是不存在的。

稍后，阿贝尔严格证明了高于四次的一般代数方程不可能有一般形式的根式解。但是这并不等于说任意一个具体数字系

数的高次方程都没有根式解。阿贝尔还是没有彻底解决代数方程的根式解问题，余下的任务是确定哪些方程可用根式求解，伽罗瓦理论正是为了解决这个问题而引起的。

受拉格朗日的影响，伽罗瓦相信方程能否有解与方程诸根的置换性质有内在联系。一个n次方程的n个根x_1，x_2，…，x_n，共有$n!$个可能的置换，它们的集合关于置换的乘法构成一个群，叫做根的置换群。方程的可解性可以在根的置换群的某些性质中反映出来。基于这一认识，伽罗瓦把方程论的问题转化为群论的问题来解决。

大家熟知，代数方程的根式解公式是具有"层次式结构"的。例如，二次方程求根公式中包含一层平方根；三次方程求根公式中有两层根号，里面一层是平方根式，外面一层是立方根式。一般说来，一个高次代数方程如果存在根式解公式，则公式中必将包含由开方根运算构成的一些层次。伽罗瓦的基本思想方法是把层次结构的形成同域的不断扩张概念联系起来；把每一层次的对应域的形成要素归结为预解式和预解方程的寻求；把预解式的寻求归结为置换群的各阶子群的结构分析。

伽罗瓦曾考虑过一类特殊方程：

$$x^4+px^2+q=0$$

这是一个可用根式求解的方程。我们从这一具体方程入手，考察可用根式求解的方程的特性，从而阐明伽罗瓦的思想方法要点。

令R_0是由p和q的有理表达式所形成的域，这些表达式的系数在有理数域中。按伽罗瓦的说法，R_0是由添加字母p、q到有理数域中而得到的域。这个域R_0叫做给定方程的系数域，而这个方程就说是属于这个域R_0。伽罗瓦并没有用域这个术语，但他确实用了这一概念。

不难看出，这四次方程的根是

$$x_1=\sqrt{\frac{-p+\sqrt{p^2-4q}}{2}},\qquad x_2=-\sqrt{\frac{-p+\sqrt{p^2-4q}}{2}},$$

$$x_3=\sqrt{\frac{-p-\sqrt{p^2-4q}}{2}},\qquad x_4=-\sqrt{\frac{-p-\sqrt{p^2-4q}}{2}}.$$

这些根满足

$$x_1+x_2=0,\qquad x_3+x_4=0$$

而这些关系式的系数在R_0中。由于给定的方程是四次的，因而存在根的24个可能的置换，它们构成一个24阶的置换群，记为S。其中有且仅有下面八个置换

$$E=\begin{pmatrix}x_1 & x_2 & x_3 & x_4\\ x_1 & x_2 & x_3 & x_4\end{pmatrix},E_1=\begin{pmatrix}x_1 & x_2 & x_3 & x_4\\ x_2 & x_1 & x_3 & x_4\end{pmatrix}$$

$$E_2=\begin{pmatrix}x_1 & x_2 & x_3 & x_4\\ x_1 & x_2 & x_4 & x_3\end{pmatrix},\ E_3=\begin{pmatrix}x_1 & x_2 & x_3 & x_4\\ x_2 & x_1 & x_4 & x_3\end{pmatrix}$$

$$E_4=\begin{pmatrix}x_1 & x_2 & x_3 & x_4\\ x_3 & x_4 & x_1 & x_2\end{pmatrix},\ E_5=\begin{pmatrix}x_1 & x_2 & x_3 & x_4\\ x_4 & x_3 & x_1 & x_2\end{pmatrix}$$

$$E_6=\begin{pmatrix}x_1 & x_2 & x_3 & x_4\\ x_3 & x_4 & x_2 & x_1\end{pmatrix},\ E_7=\begin{pmatrix}x_1 & x_2 & x_3 & x_4\\ x_4 & x_3 & x_2 & x_1\end{pmatrix}$$

使得上述系数在R_0中的两个关系式保持不变。这八个置换构成一个群S的子群，记为H_0，H_0称为给定方程相对于域R_0的群。就是说，一个方程相对于域R_0的群是根的置换群或子群，这群中的置换使给定方程的根之间带有R_0中的系数的全部关系不变。H_0中置换的数目是我们对根的无知程度的一个尺度，因为在这八个置换之下我们不能把它们区分开来，

所以H_0的阶实际上就表示出根的不可区分的程度。为了解出各个根x_k，就必须对它们进行区分。

现在考虑$x_1^2-x_3^2$，它等于$\sqrt{p^2-4q}$。添加这个根式到R_0中，形成一个域R_1，即形成包含R_0和$\sqrt{p^2-4q}$的最小的域。于是

$$x_1^2-x_3^2=\sqrt{p^2-4q}$$

是R_1中的一个关系。考虑到$x_1^2=x_2^2$和$x_3^2=x_4^2$（因为$x_1+x_2=0$和$x_3+x_4=0$），我们可以看出，群H_0中前四个置换使R_1中这个关系保持不变，但后四个不行。这前四个置换构成H_0的一个最大不变子群，记为H_1。H_1就是方程相对于R_1的群。

现设我们添加量$\sqrt{(-p-D)/2}$到R_1中（这里，$D=\sqrt{p^2-4q}$）并形成域R_2，则

$$x_3-x_4=2\sqrt{\frac{-p-D}{2}}$$

是R_2中的一个关系。这个关系仅在头两个置换E和E_1下保持不变，而在其余六个置换下不这样。这两个置换构成H_1的一个最大不变子群，记为H_2，H_2就是方程相对于R_2的群。

假设添加量$\sqrt{(-p+D)/2}$到R_2中，又得到域R_3。在R_3中我们有

$$x_1-x_2=2\sqrt{\frac{-p+D}{2}}$$

现在恰有置换E是使R_3中全部关系保持正确的仅有置换，它构成单元素群H_3，H_3是H_2的最大不变子群。它是方程相对于R_3的群。

伽罗瓦证明了，当方程相对于某个域的群恰巧只含恒等置

换E时，这个域就是该方程的根域(包含方程所有根的最小域)。因此，R_3是给定方程的根域。就是说，用这域中元素就足以把方程诸根明确表示出来。

一个方程可否用根式求解，与根域的性质密切相关。如果从系数域到根域的扩域过程中每次添加的都是根式，则方程可用根式解，否则，方程不能用根式解。上述三次扩域所添加的都是根式，都取自原方程的根的表达式中。对于一般的高次方程，需要的是在不知道根的情况下判别它可否用根式求解。为了确定扩域过程中的添加项，伽罗瓦也使用预解式概念，给出构造预解式的方法。仍就原方程来讨论。根据伽罗瓦的方法，为扩充域R_0，需要在R_0中构造一个方程，叫做原方程的部分预解式。这个预解式就是$t^2-(p^2-4q)=0$，它的次数等于子群H_1在母群H_0中的指数(母群的阶被子群的阶除所得的商)$8\div4=2$。这个预解式的根，例如，$\varphi_1=\sqrt{p_2-4q}$，就是为扩充域R_0所要的添加项。再在域R_1中构造方程的预解式

$$t^2-2(-p-\sqrt{p^2-4q})=0$$

它的次数等于子群H_2在母群H_1中的指数$4\div2=2$。解这个预解式，取其根

$$\varphi_2=\sqrt{2(-p-\sqrt{p^2-4q})}$$

为扩充域R_1所要的添加项。同理，在R_2中作方程的预解式

$$t^2-2(-p+\sqrt{p^2-4q})=0$$

它的次数等于H_3在H_2中的指数$2\div1=2$。解之得

$$\varphi_3=\sqrt{2(-p+\sqrt{p^2-4q})}$$

就是扩充域R_2所需添加项。根据伽罗瓦理论，上述步骤对于一般的代数方程也是可行的。如果扩域过程中任一预解式都是

二项方程

$$x^k - A = 0 \qquad (k为质数)$$

则预解式可用根式解，从系数域到根域的扩域过程所添加的都是根式，方程便是可用根式求解的。反之，如果一个方程可用根式求解，则预解式方程组必定存在，并且都是次数为素数的二项方程。

上述讨论涉及两个序列。一个是扩域序列

$$R_0 \subset R_1 \subset R_2 \subset R_3$$

另一个是最大不变子群序列

$$H_0 \supset H_1 \supset H_2 \supset H_3$$

其中，H_i是方程相对于域R_i的群。后一序列又叫合成序列。对于一个给定的方程，寻找相对于系数域的群及最大不变子群序列，纯粹是群论的事，是可以做到的。有了这个子群序列，子群H_{i+1}在母群H_i中的指数也随之确定。伽罗瓦得到一个重要准则（伽罗瓦基本定理）：如果这些指数都是素数，则方程可以用根式求解；若这些指数不全是素数，则方程不可用根式求解。这样，方程的可解性问题，这一历史难题，就可完全用群论的方法去解决。

伽罗瓦引用了一系列新名称、新符号，他的理论太深了，远远走在同代人的前面，的确“使读者产生反感”。他的论文时而被数学大师“遗失”。时而又被数学大师认为“难以理解”，直到他死后多年才得以发表，为人们所理解和接受。

伽罗瓦引用了大量的新名称、新符号是“因为题目的重要性”，是值得的。他为人类作出了杰出的贡献，解决了困扰数学家数百年之久的问题。更为重要的是，通过方程可解性问题，他提出了全新的数学概念，开辟了全新的研究领域，对近世代

数的形成和发展产生了巨大影响。

如果没有多大的必要性而引用过多的符号，就是“丑陋的”。正如J·模棣尔所说，丑陋的数学是“用了那么大量的变量，常数和上、下、左、右标，使得人们很难掌握结果的含义，他显然只是为其自身需要进行推广，产生的结果也没什么新颖之处。另外还有一些工作，它们的结果没有显著的重要性，因此也不应该在理解和验证的方式上给读者加重负担。”

第五章 “问题解决”中的符号处理技巧

符号学为人们提供了一个新的了解、分析研究问题的思维方法。人类几乎处处都要运用符号处理技术。利用计算机解决问题，要对问题作符号处理，利用电报传递信息要对电文作符号处理……在数学研究中表现得尤为著目，我们熟知，韦达引用符号促进代数学成为独立学科，莱布尼茨引用符号创立了微积分，伽罗瓦引用符号创立了群论……现代，“符号处理”已经发展成“问题解决”过程中的重要方法。

§1 问题与“问题解决”

许多自然科学家、心理学家认为，科学研究与创造是“从提出问题开始的”。爱因斯坦曾明确说过，提出一个问题往往比解决一个问题更重要。英国科学家波普则把科学发现的逻辑表述为“$P_1 \longrightarrow TT \longrightarrow EE \longrightarrow P_2$”,即“问题$_1$ ⟶试探性理论⟶消除错误⟶问题$_2$”这样一个图式，认为“科学只能发端于问题”。苏联心理学家鲁宾斯坦也早就说过，“思维起源于问题的情景”，“是从分析问题情景开始的”。

一、数学问题与数学发展

什么叫问题呢？心理学家认为，所谓问题实际上就是出现了“没有加以确定的环节或者要求的情景，这种情景是以其中揭

示出的某种东西为前提的”。问题与习题有区别，例如，解一元二次方程对高中生就不是问题，而对初一学生就成问题了。一般说来，一个“情景”可以产生一个问题，也可以不产生什么问题。一个情景的要求，倘若毫无困难地通过一些明显的行动就达到了所求的目标，那就不产生问题。然而，倘若谁想不出这样的行动来，那就产生了问题。人们在日常生活、科学活动中经常会发现问题。

美国数学家哈尔莫斯曾指出：定理、证明、概念、定义、理论、公式、方法中的任何一个都不是数学的心脏，只有问题是数学的心脏。数学科学的起源和发展是由问题引起的。古人结绳计数是为了解决生产和生活用品的多少问题。我们秦汉时期的数学著作，《周髀算经》和《九章算术》就是当时的数学家解决数学应用问题成果的汇集。几何学萌芽于古埃及，那是从解决尼罗河流域的土地测量问题而产生，后来经过古希腊过滤、发展起来的。诸如自古相传的立方倍积问题、化圆为方问题、三等分角问题；数值方程的解、曲线论、微积分、傅里叶级数和位势理论中那些最初的问题；还有大量的属于力学、天文和物理学方面的问题等等，都刺激、推动了数学的发展。

自希尔伯特1900年在巴黎会议上发表讲演《数学问题》以来，数学问题对于数学发展的重要意义已经得到公认。正如希尔伯特所说：“某类问题对于一般数学进展的深远意义以及它们在研究者个人的工作中所起的重要作用是不可否认的。只要一门科学分支能提出大量的问题，它就充满着生命力；而问题缺乏则预示着独立发展的衰亡或中止。”历史事实表明，通过提出问题会导致整门新学科的诞生。希尔伯特举了三个典型例子。第一，贝努利最速下降线问题是现代数学分支——变分法

的起源。第二，费马问题，它看上去“非常特殊，似乎不十分重要”，却对科学产生了令人鼓舞的影响。现代数论中的核心概念“理想数”正是为了解决费马问题而提出的，而且其意义已远远超出数论的范围而深入到代数和函数论的领域。第三，三体问题，它对现代天体力学起了关键作用。

当然，类似的例子还有不少。例如，分赌注问题与概率论，七桥问题与拓扑等。

希尔伯特悟得一条真理：重大的个别问题是数学的活的血液。为了说明某些问题的重要性，希尔伯特还在《数学问题》讲演中提到了维尔斯特拉斯：

“维尔斯特拉斯认为他的极大的幸运是在其科学事业之初，就找到了像雅可比逆问题这样一个重要的、可供研究的问题。”

希尔伯特的演讲《数学问题》是世界数学史的重要里程碑。

希尔伯特高瞻远瞩地用23个数学问题，预示20世纪数学发展的进程。巴黎会议之后，各国数学杂志纷纷转载他的演说稿，大批数学家投入解决“23个问题”的激流中去。1900年希尔伯特的学生，22岁的麦克思·戴恩获得第一个重要结果。他证明了(正如希尔伯特所猜想的那样)一个正四面体不可能被剖分以后再拼合成一个体积相等的立方体。这给出了希尔伯特第3问题的部分解答。次年，戴恩完全地证明了第3问题——两个等底等高四面体的体积相等问题。其大意是：存在两个等底等高的四面体，它们不可能分解为有限个小四面体，使这两组四面体彼此全等。戴恩证明确实存在着这样的两个四面体。后来，他被称为“荣誉等级”的数学家行列中的第一名数学家。

现在，时光已过去80多年，“23个问题”约有一半已获得解决，有一些取得了很大进展，有些则收效甚微。1975年，在美国的伊利诺斯大学召开了一次国际数学会议，邀请世界著名数学家参加，专门研究“23个问题”的进展。会后出版的论文集详细介绍了各个问题的进展。80年代以来，人们把解决希尔伯特问题，那怕是其中一部分，都看成至高无上的荣誉。据统计，从1936年到1974年，菲尔茨国际数学奖的20名获奖人员中，至少有12人的工作与希尔伯特问题有关。1976年，美国数学会组织评论1940年以来的美国十大数学成就，其中有3项是希尔伯特问题的解决。我国数学家陈景润在希尔伯特第8问题（素数问题）上取得了世界领先地位，在第16问题上也作出了一些贡献。

大数学家韦尔在希尔伯特去世时的悼词中曾说过：“希尔伯特就像穿杂色衣服的风笛手，他那甜蜜的笛声诱惑了如此众多的老鼠，跟着他跳进了数学的深河。”希尔伯特提出“23个问题”之后，果然形成了许多新的数学分支，达到了他预期的目的。

数学问题与数学发展之间的密切关系，由此可见一斑。

二、“问题解决”

日本心理学家大桥正夫曾讲过，思维是“解决问题的过程”。问题的解决是一种高级形式的智力活动。它是指人们在社会实践和理论学习中，面临新情景、新课题，而又没有现成对策、答案或解决方法时，所引起的寻求处理问题的一种紧张心理活动。这种心理活动，具有某种程度的创造性。

由于问题的性质不同以及主体思维方式相异，因此问题解

决的心理过程是多种多样的。不可能用一个统一的模式来解释如此复杂多样的智力活动过程，许多心理学家和教育家从不同的角度对之进行了描述，企图将问题解决的过程清晰地呈现出来。在这方面，1945年G·波里亚的《怎样解题》被誉为一面"问题解决"的旗帜。G·波里亚认为，面对问题，解决它，就"意味着发现一条摆脱疑难，绕过障碍的途径，以达到一个不能一蹴而就的目的"，为此需要动员和组织已有的知识、技巧，"它是人类最富有特征的智力活动"。《怎样解题》的中心思想就是谈问题解决过程中怎样诱发灵感。其中的"怎样解题"表实质上就是试图用以诱发灵感的"智力活动"表。"表"中包含四部分内容：弄清问题；拟定计划；实现计划；回顾。G·波里亚写道："了解问题是为了好念头的出现作准备；制订计划是试图引发它；在引发之后，我们实现它；回顾此过程和求解的结果，是试图更好的利用它。"其中的所谓好念头就是指数学灵感。

在数学学科中，能力指的是什么？G·波里亚说："这就是解决问题的才智——我们这里所指的问题，不仅仅是寻常的，它们还要求人们具有某种程度的独立见解、判断力、能动性和创造精神。"因此，他提出："中学数学教学首要的任务就是加强解题的训练。"

G·波里亚发现，在日常解题和为攻克难关而作出的数学中的重大发现之间，并没有不可逾越的鸿沟。他写道：

"一个重大的发现可以解决一个重大的问题，但在求解任何问题的过程中，也都含有点滴的发现。"

"一个有意义的题目的求解，为解此题所花的努力和由此得到的见解，可以打开通向一门新科学，甚至通向一个科学新

纪元的门户。”

G·波里亚认为，要想作出重大的数学发现，就必须重视平时的解题，因为平时的解题和数学发现之间，只有难易程度上的差别，在本质上是完全一样的。

G·波里亚的“解题”不同于通常人们所指责的“题海战术”。他认为，一个数学教师，如果“把分配给他的时间塞满了例行运算来训练他的学生，他就扼杀了学生的兴趣，妨碍了他们的智力发展……”。因此他主张，与其穷于应付繁琐的数学内容和过量的题目，还不如选择一个有意义的但又不太复杂的题目，去帮助学生深入发掘题目的各个侧面，使学生通过这道题目，就如同通过一道大门而进入一个崭新的天地。比如，“证明$\sqrt{2}$是无理数”和“证明素数有无限多个”就是这样的好题目，前者通向实数的精确概念，而后者是通向数论的门户。打开数学发现大门的金钥匙往往就在这类题目之中。

众所周知，数学有两个侧面，一方面，已严格提出来的数学是一门系统的演译科学，另一方面，在创造过程中的数学看来却像一门实验性的归纳科学。G·波里亚指出，通过研究解题的方法，我们可以看到数学的第二个侧面，也就是看到“处于发现过程中的数学”。这对于提高创造能力是有益的。

《怎样解题》已经译成16种文字发行100多万册，成为世界名著。著名的现代数学家瓦尔登在1952年2月2日瑞士苏黎世大学的会议致词中就曾说过：“每个大学生、每个学者，特别是每个教师都应该读这本引人入胜的书”。对这本书给予极高的评价。今天，人们已公认，在“问题解决”的研究方面，G·波里亚是作出了划时代的贡献。

G·波里亚把“解题”作为培养学生的数学才能，教会他

们思考的一种手段和途径。这种思想得到了国际教育界的广泛赞同，1976年国际数学管理者委员会把解题能力列为十项基本技能的首位。1980年，美国数学教师全国委员会的“行动纲领”建议把“问题解决作为80年代学校数学的核心”，这是继“新数学运动”和“回到基础”之后的又一新口号。“问题解决”已成为美国数学教育工作者的目标，而G·波里亚则成为“问题解决”的带头人。目前，以“问题解决”为中心的课程改革，正在世界许多国家探讨、研究。人们普遍认为，这种“问题解决”的数学课程，有利于锻炼学生解决问题的能力，有利于培养现代社会所需要的，具有创造能力的人才。这种数学课程，只要再注意到数学知识的系统学习，就能比较符合社会结构、数学知识结构和学生心理结构发展的要求。

§2 几何图示法

笛卡尔曾说过：“没有任何东西比几何图形更容易印入脑际了，因此用这种方式来表达事物是非常有益的。”G·波里亚正是用几何中的线来表示思维过程中推理的线索而得到了“解题过程的几何图示法”，这种几何图示法就是一种利用符号处理问题的方法。

G·波里亚回忆说，在他读大学的时候帮一个男孩子复习功课，被一条立体几何题“卡住了”。“居然对付不了这样一道简单的题目，我只能责怪自己。第二天晚上我坐下来重新从头到尾仔细去解它，那次我是做得这样彻底，以致这辈子再也不会忘记它了。由于试图直观地去看清楚整个解题的自然过程和解题中涉及的一系列基本想法，我终于得到了一个解题过程

的几何图示法。”G·波里亚写道：“这是我对解题的第一个发现，也是我终生对解题产生兴趣的开端”。

发现解法，就是在原先是隔开的事物或想法（已有的事物和要求的事物，已知量和未知量，假设和结论）之间去找出联系。被联系的事物原来离得越远，联系的发现者的功绩也就越大。这种联系就像一座桥：一个伟大的发现使我们强烈地觉得像是在两个离得很远的想法的鸿沟上方架上了桥。我们常常看到这种联系是由一条链来贯穿的：一个证明像是一串论据，像是一条由一系列结论组成的链，也许是一条长链。对于思维上的联系人们经常使用的词是线索，将一条细微的线索当成一条几何上的线，将被联系着的事物当成几何上的点，这样无可避免地，一幅隐喻着一系列数学结论的图式便必然地浮现出来了。

G·波里亚选择了一个非常简单的立体几何题，来阐明他的方法，并把一系列引导到它的证明的想法重新整理出来。他“缓慢地，非常缓慢地去作，逐个地一个接一个地把线索揭示出来”。从中我们可以看出他是怎样随着他的思想的发展而逐步引入符号的。

问题：给定正棱台的高h，上底与下底分别是边长为a与b的正方形，求正棱台的体积V。

一、明确问题与目标，并用符号表示

先集中到目标上，并尽可能明确地画出所求的体积$\overline{V}$的图形（见图5－1的左侧）。所求目标在思维中的位置则用一个单点$\overline{V}$，象征性地表示出来，我们的全部注意力应该集中在它上面（见图5－1的右侧）。

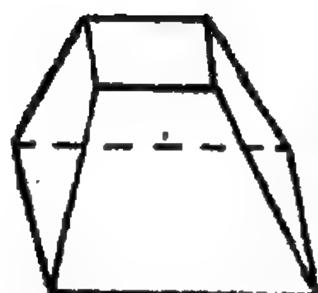


图5-1　集中到一点：目标

我们问自己：已知量是什么？这样，我们就把注意力集中到图中那些长度已定的线段a、b和h上，见图5-2的左侧。

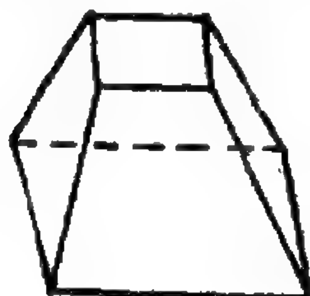


图5-2　未定的问题：沟上架桥

我们在图5-2右侧增加三个新点，分别记作a、h和b，来代表三个已知量在变化了的思维中的位置，它们与$\overline{V}$之间有一道鸿沟，这就是图5-2右边的那一片空白。这片空白象征着尚未解决的问题：我们的问题现在就集中为将未知量$\overline{V}$与已知量a、h和b联系起来，我们必须在它们之间的那道鸿沟上架起桥来。

二、转化问题，引进新符号

工作的最初阶段已由图5-1和图5-2适当地描绘出来了。下一步该怎么走呢？

当我们不能解出新提出的问题的时候，总可设法寻找一个适当的有关联的问题。眼下，情况比较简单、未知量是什么？一个棱台的体积。这个棱台是我们用一个平行于底的平面，截去整个棱锥中的一个较小的棱锥以后所剩下的部分。大（整个）棱锥的底是一个面积为b^2的正方形，见图5-3。如果我们知道这两个棱锥的体积B和A，我们就能求出棱台的体积：

$\overline{V}=B-A$

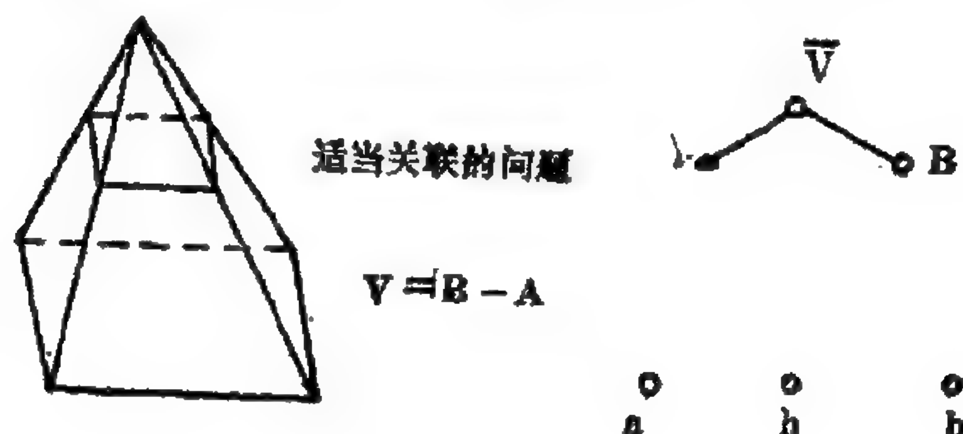


图5-3

如果你不能解所出的问题，那就去寻找一个……

也许是一个好主意！

于是我们原来的问题——求$\overline{V}$，就转变成了两个适当关联着的辅助问题——求A和B。为了用图式来表达这一转变，我们在未知量$\overline{V}$和已知量a，h，b中间的那片空白上引进两个新

的点，记为A和B。我们用斜线把A和B与$\overline{V}$联结起来，以此来表示这三个量之间的基本关系：从A和B出发，我们就能得到$\overline{V}$，即关于$\overline{V}$的问题的解法是建立在关于A和B这两个问题的解法的基础上的。

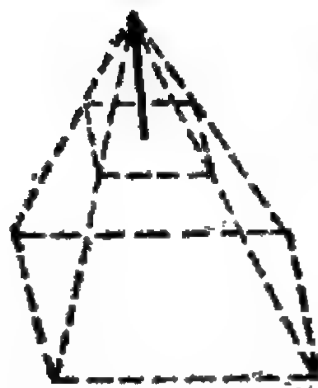
下面的任务是要去求两个新的未知量A和B。在图5－3中，两个未决的点A和B与已知量a、h和b之间还隔着一道鸿沟。不过事情看来是有希望的，因为棱锥对我们来说是比棱台更熟悉的图形，而且，虽然现在替代未知量$\overline{V}$的是两个未知量A和B，但这两个未知量是完全类似的，而且分别和已知量a和b有类似的关系。对应地，在图5－3中，思维状态的图式表示是对称的。我们已经开始在原来的未知量和已知量之间那片空白上架桥了；剩下的部分就窄多了。

三、进一步发展

我们要求找出未知量A和B。未知量A是什么？一个棱锥的体积。如果我们有两个已知量：底面积和棱锥的高，这个棱锥的体积就能算出来。这里高度并没有给出来，不过我们还是可以把它考虑进来，设它是x，则

$$A=\frac{a^2x}{3}$$

在图5－4的左侧，棱台上面的小棱锥显示得较为明白，我们特别强调了它的高x。这一阶段的工作则在图5－4的右侧给出了图式表示，在已知量的上面出现了一个新的点x，用斜线分别把A和x，A和a联结起来，表示由x和a能得出A，即A能用x和a表示出来。虽然还有不定的点悬在空中和已知量的



你怎样才能得到
这样一类的量？

$$A=\frac{a^2X}{3}$$



图 5－4

第一个联系已作出，但仍有两个未知量（在图 5－4 中尚悬在半空的不确定的点）需要去求，但我们已跨出了前进的一步，因为我们至少已经成功地把未知量V和一个已知量a联系起来了。

下一个步骤是显然的。未知量A和B具有类似的性质（它们在图 5－3 中是对称地表示出来的）。我们已经把体积A用底和高表示出来了，我们也能把体积B类似地表示为

$$B=\frac{b^2(x+h)}{3}$$

在图 5－5 的左侧部分，含棱台的大棱锥显示得比较明白，这里特别强调了它的高$x+h$。在图 5－5 的右侧部分出现了三条新的斜线；把B分别与b、h和x表示出来，于是只有一个点还悬着——点x还没有与已知量联结起来。沟变得更加狭窄，它现在只横在x和已知量当中了。

剩下的未知量是什么？是x——一条线段的长。

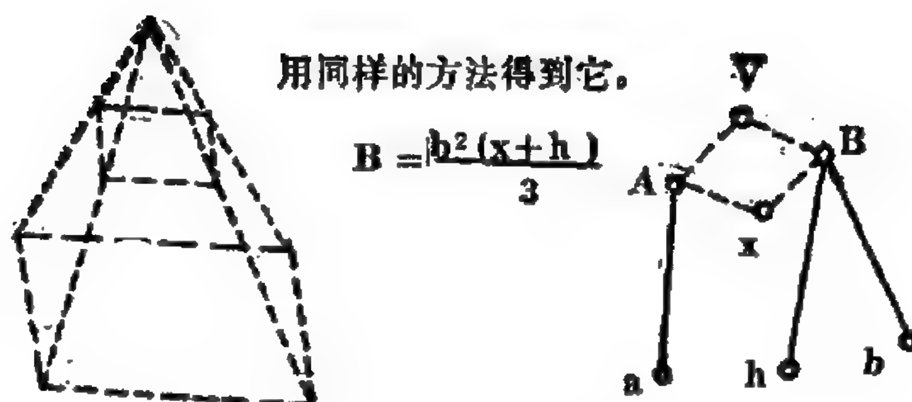


图 5-5　只剩下一个问题悬而未决了

几何中最常见的事就是从一个三角形——如果有可能就从一个直角三角形，或是从一对相似三角形去得出一条线段的长。然而图形里还没有一个用得上的三角形，而这里应该有一个以x为边的三角形。这样一个三角形应当在一个通过体积为A的小棱锥的高的平面上，这个平面同时也应当通过体积为B的大棱锥的高，而大棱锥与小棱锥是相似的。是的，通过这个高并且与棱台底的一条给定边平行的平面上有两个相似三角形。这就是我们所要的！好了，完成了！

图 5-6 中显示了一对相似三角形，由此，x 可以很容易地通过下列比例式算出：

$$\frac{x}{x+h}=\frac{a}{b}$$

细节在这个阶段并不重要，重要的乃是现在x能够由三个已知量a、h和b表示出来了。图 5-6 右侧部分出现的三条新的斜线正好表示了x可以与a、h和b联系起来。

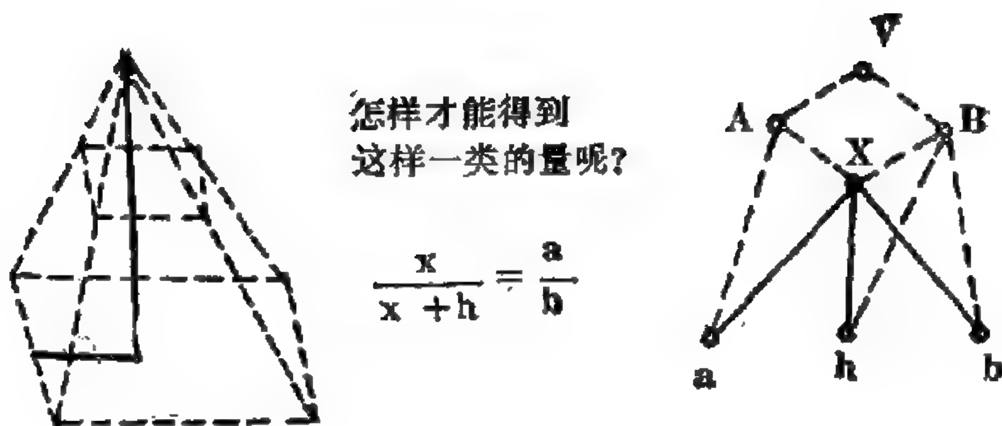


图 5－6　我们已经成功地在鸿沟上架起了桥

好！我们已经成功地在鸿沟上架起了桥，成功地通过中间量（辅助未知量）A、B和x，在未知量$\overline{V}$和已知量a、h和b之间建立起一个不中断的联络网。

四、彻底完成它

问题解决了吗？我们应当把棱台的体积V用已知量a、h和b表示出来，而这一点还没有做到。但我们工作中比较重要也是比较吸引人的部分已经过去了；剩下的任务只是些按部就班的事，不再需要拐什么弯了。

在上述工作中是有冒险的因素的。在每一个阶段我们总是希望下一步会使我们更靠近我们的目标——在鸿沟上架起桥来。当然，我们希望是这样，但我们并不十分有把握。在每一阶段我们都必须把下一步想出来并且冒险地去试它。但现在再也不需要什么发明和冒险了。我们已经预见到只要沿着图5－6中那个不中断的联络网中的线索走去，就能够万无一失地从

已知量a、h和b到达未知量$\overline{V}$。

我们下面的工作就从上述工作的结尾处开始，首先来处理最后引进的辅助未知量x。由第三段中最后一个等式可得

$$x=\frac{ah}{b-a}$$

然后把x的值代入第三段所得到的两个等式中，得

$$A=\frac{a^3h}{3(b-a)},\qquad B=\frac{b^3h}{3(b-a)}$$

这两个结果之间的类似之处使人感到欣慰，它完全符合我们的直觉。最后我们就可得到等式

$$\overline{V}=B-A=\frac{b^3-a^3}{b-a}\cdot\frac{h}{3}$$

$$\overline{V}=\frac{a^2+ab+b^2}{3}h$$

这就是所要求的表达式。

图5-7恰当地把这一部分工作用符号描述了出来，其中

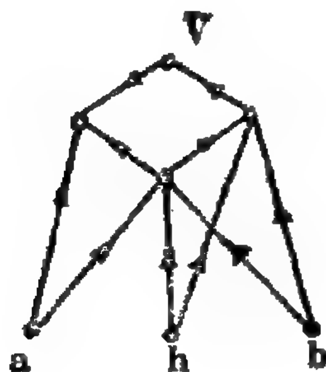


图5-7　从已知量到未知量的工作图

每一条联线都带有箭头，它指出我们是朝哪个方向去从事这一联系的。我们从已知量a、h和b出发，由此向前经过中间的辅助未知量x，A和B到达原来的、主要的未知量V，通过已知量一个一个地把这些量表示出来。

§3 心智图象

心智图象是“问题解决”过程中的深层次的符号。

心智图象是具有某种程度抽象的、模式化了的模糊“形象”，它又称为“心理意象”、“智力图象”。这种心智图象不是一般意义上的具体图形，而是一种形式化的抽象，但是，它又与词语的抽象不同，与代数的抽象不同，还保留某种图象。例如，把一条奔腾咆哮的大江抽象为一条曲线（不是抽象为“江”这个词），把两个有公共元素的学生集合抽象为两个相交的圆。欧拉为了向一个瑞典王子解释演绎的特性，就是用圆来代表一般的概念。让我们考虑A和B两类事物，如果“凡是A都是B”，则我们就想象圆A位于圆B之内；而如果“没有A是B”，我们就想象圆A和圆B完全不相交，如果“某些A是B”则想象成圆A和圆B相交。

例 有41名学生参加数理化三科竞赛，其中不及格的人数为：

数学	物理	化学	数理	数化	理化	数理化
12	5	8	2	6	3	1

试问有多少学生三科都及格？

如图5－8所示，不妨先画出两两相交并有公共部分的三

个圆，再顺次由数理化、数理、数化、理化不及格的人数，分别填好数据。因总人数为41，显然，三科都及格的人数为41－15＝26（人）。

这样，借助于心智图象，就完成了本题的求解。

构作心智图象是对于问题进行符号处理的有效方法。正如笛卡尔所说："在用推理解决问题时，心智图象的作用是首要的。"G·波里亚的几何图示法实际上就是构作心智图象。

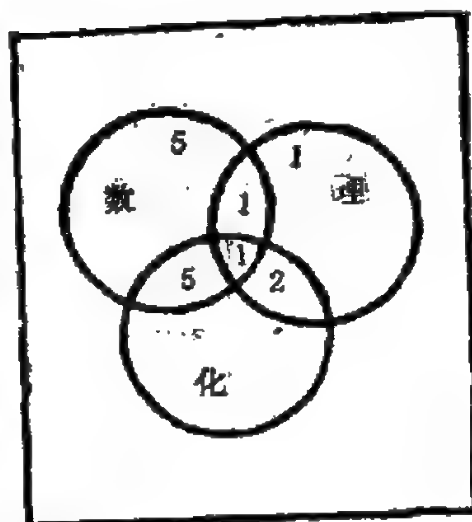


图5－8

一、心智图象的整体识别功能

心智图象具有综合的特点，利用心智图象便于从整体上进行直觉的识别。

美国学者西蒙曾经介绍过这样一个实验：在棋手面前将一次高水平棋赛的某个棋谱（含有25个棋子）展示10秒钟，然后要求受试者凭记忆复述这个棋谱，如果是国际象棋大师，能准确无误地全部复述出来；如果是新手，只能复述约6个子。倘若同样数量的棋子是在棋盘上随意乱放的，那么，结果是象棋大师与新手一样只能复述出6个子。造成这种情形的原因在于象棋大师对棋谱是一种整体识别，他建立了某种心智图象——一个具有内在联系的棋谱。10秒钟的注视使他把储存在脑子里的心智图象整块地取了出来，而不是一只只棋子的记忆。一旦遇到杂乱无章的棋子，他就搜索不到已有的心智图象，他那长期形成

的模式就不起作用了。只能跟新手一样死记。

这种从整体上构筑心智图象的能力是现代电子计算机所缺乏的。十年前，美国最好的电子计算机棋手Belle与世界冠军相比，评分为1.90:2.70分，相差甚多。尤其在残局时Belle的能力尤低，因为机器的长处在于逻辑运算迅速，缺点在于它不能从全局通盘地估计形势；而人的优点则在于认出形势，人能借助于心智图象从整体上进行直觉的判别。

笛卡尔高度重视心智图象在数学推理中的整体把握作用，他说："在把推理过程中的结果一一罗列之后，就需要记住它们，而记忆可以帮助我们把那些暂时不用的资料贮存起来。但若这些被考虑的资料既不按心智图象的方式经常在脑海中出现，又不将它们在各个例子中全部奉献出来，那么，这些资料就有被忘掉的危险。"事实上，为了从整体上把握研究的对象与方向，很多数学家、科学家都有构建心智图象的习惯。希尔伯特的多维空间不是现实的形象空间，它是思维所构想的心智图象空间。爱因斯坦相对论的四维空间世界，原子结构理论中的"玻尔轨道"、"电子壳层"、"电子云"等等都是一种具有某种抽象性的心智图象模型。这里并非真有形象性的"轨道"、"云"之类的东西。

二、关于"素数个数无限性"的证明

关于素数，在历史上最初所产生的一个问题是：素数是有限多个还是无限多个？欧几里德首先证明了："素数的个数是无限的"。这是算术中的一个基本而著名的定理。

欧几里德没有拘泥于素数的细节特征而是从整体上把握了这个问题的研究内容和方向：如果素数的总体是有限个，能不

能推到矛盾？设有最大素数P，则$2\times3\times5\times\cdots\times P+1$是素数还是合数？……这是导致成功的方向。

阿达玛依次列出了这个经典证明中的各个步骤，同时描述了他在“读到这个证明的每一步时的心智图象”（假设，我们要证明存在着比11大的素数）：

证明步骤	心智图象
（1）考虑所有从2到11的素数，即2，3，5，7，11	（1）我看到一堆乱七八糟的数
（2）作乘积 $2\times3\times5\times7\times11=Q$	（2）Q是一个相当大的数，我眼前出现一个点，它远离那堆乱七八糟的数
（3）在这个乘积上加1	（3）我见到了第二个点，稍稍离开第一个点
（4）此数若不是素数，必定能被一个素数除尽，这个素数就是所求的数	（4）我在那一堆数和第一个点之间看到一个位置

阿达玛说，这种奇怪而又模糊的符号对于弄懂上面的证明是十分重要的，因为，借助于它，“我就可以一下子看到论证中的所有成分，把它们相互联结起来，并使之成为一个整体——一句话，达到综合的目的。”

利用智力图象便于从整体上把握研究的对象与方向，这是一种有效的符号法。阿达玛说，在他所从事的全部数学研究中，他都会构作心智图象。

§4　符号处理技巧举例

几乎每个数学分支都有一套符号处理技术。对于一些特殊问题，除了构作心智图象外，往往还需要一些特殊的符号处理技巧。在本节中，我们将通过一些较为典型的实例来进一步阐明“问题解决”过程中的符号处理技巧。其中有些例子，完全可以采用某些现代数学的原理来解决，不过，为了揭示符号处理技巧，为了欣赏“处于发现过程中的数学”，我们宁可撇开现成的原理，而引用符号解决问题。

一、哥尼斯堡七桥问题

这是数学史上有名的例子，几乎每一本数学方法论的著作都要谈到它，我们当然也不能放过它。

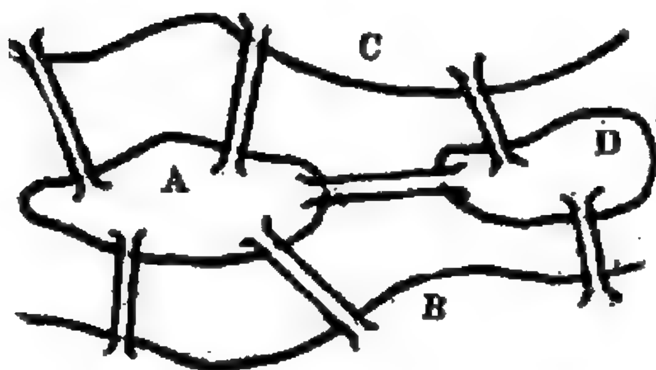


图5-9

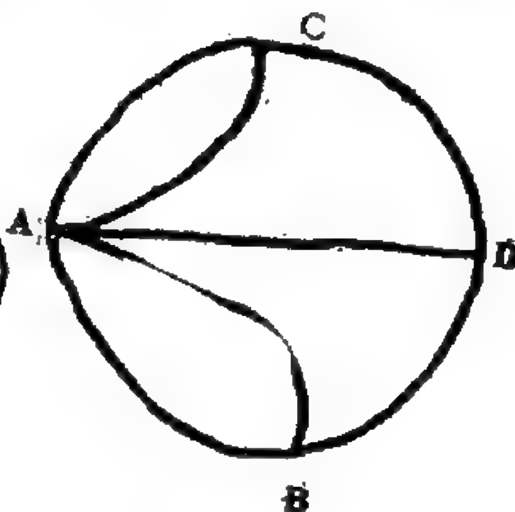


图5-10

布勒格尔河流经哥尼斯堡城市区，河中有两个小岛（如图5-9中A、D），小岛与河岸有七桥相连。人们长期思考一个

有趣的问题：一个游人能不能不重复、也不遗漏地一次走遍这七座桥而回到原出发地？欧拉在1736年解决了这个问题，并在彼得堡科学院作了一个科学报告。

欧拉看出小岛、河岸的大小、形状以及桥的长度都无关紧要，而相互之间的连接关系才是问题的实质。因此，他就把小岛A、D与两岸B、C都用相应的符号——“点”来表示。而连接两地的桥也用符号来表示，表示为连接两“点”的弧线（如图5-10，这个心智图象画成什么样子是次要的，只要不改变元素之间的结合关系）。为此，通过一座桥就可看作画一条弧，而不重复地通过七座桥就是要一笔画出七条弧所组成的图形。这样问题就转化为“一笔画”问题了。

欧拉考虑到“一笔画”的结构特征。一笔画有一个起点和一个终点，除起点和终点外，一笔画中可能出现一些曲线的交点，在这些交点处通过的曲线一进一出，总是偶数条，可称为“偶点”。可见，只有起点和终点处通过的曲线可能是奇数条，称为“奇点”。任何一个一笔画图形或者没有奇点，或者有两个奇点。而现在这个图形中四个点都是奇点，因此超出一笔画的范围，不能一笔画成，这意味着“七桥”不可能一次无重复地走完。

欧拉解决这个问题的关键是运用抽象思维，揭示了七桥问题的实质并对问题进行了符号处理。他的这一项成功研究对于几何学的发展起了重要作用。正如欧拉在他的科学报告中所说：

“讨论长短大小的几何分支一直被人们热心地研究着，但是，还有一个至今几乎完全没有探索过的分支，莱布尼茨最先提起过它，叫做‘位置几何学’，这个几何分支只讨论与位置有关

的关系，只研究位置的性质，它不去考虑长短大小，也不牵涉到量的计算。但是至今未有过令人满意的定义，用以刻划这门位置几何学的课题与方法。近来流传着一个问题，它虽然无疑是属于几何学的，但不是求一个尺寸，也不能用量的计算来解答，所以我毫不犹豫地把它归入位置几何学，特别还因为要解答它只要考虑位置，不用计算。在这里我要讲一讲我所发现的解答这类问题的方法，可以作为位置几何学的一个例子"。

其中所说的问题就是上述七桥问题。事实上，欧拉的研究最终导致了线路拓扑学这一分支的创立。

二、拉姆齐问题

1947年，在匈牙利举行数学竞赛的时候，出了这么一道题：

证明：在任何6个人中，总有3个人相互认识或者互不认识。

这道题一出现，立即引起各国从事数学竞赛的数学教育家以及组合数学（特别是图论）专家的注目。1953年，美国普特南数学竞赛时，又出了这道题。1955年，加拿大著名组合数学专家格林伍德与格里逊专门撰写论文，对这道题作了推广，1958年，美国著名数学杂志《美国数学月刊》将这道题列为E_{1321}，广泛征求解答。以后，这道题的种种推广就经常被各种水平的竞赛用做试题。

为了证明这问题，我们先对问题进行符号处理：把6个人抽象为平面上的六个点，记为A_1，A_2，A_3，A_4，A_5，A_6。把互相认识这个关系形式化地用两点之间的实线段来表示，而不认识则用虚线段表示。这时，以A_1为出发点，向其余五点共可

连5线，依抽屉原则，至少有3条线段同类，不妨设为实线段（即互相认识），它们分别为A_1A_2，A_1A_3，A_1A_4。这时考察$\triangle A_2A_3A_4$，若它的三条边皆为虚线，结论成立；若它的三条边中至少有一条为实线段，不妨设为A_2A_3，则$\triangle A_1A_2A_3$的边为同类，结论仍然成立。

这道数学竞赛题引起如此广泛的重视，有其深刻的原因。简要地讲，这主要是因为它以组合数学的一个分支——图论作为背景的。而它本身就是组合数学中一类典型问题——拉姆齐问题的雏形。

什么叫拉姆齐问题？为了说明这问题，我们再看一个例子：

17个科学家中的每一个和其他的人都通信。在他们的通信中仅仅讨论三个题目。证明其中至少有3个科学家，他们互相通信中讨论的是同一个题目。

证明方法与上例相仿。

把科学家用点A_0，A_1，……A_{16}表示。把科学家之间互相通信讨论题目的关系表示为两点之间的连线段。为了区分不同的题目，还必须对这些线段附加不同的标志，这时因为有三个题目，用虚线、实线已不足以区分，我们用涂色的办法以示区别：如果讨论的是第一个题目，这线段涂上红色；讨论的是第二个题目，涂黄色；第三个题目，涂蓝色。

自A_0引出的线有16条，根据抽屉原则，其中必有六条或更多条是同一种颜色的。不妨设A_0A_1，A_0A_2，A_0A_3，A_0A_4，A_0A_5，A_0A_6为红色。

考虑自A_1引出的五条线段A_1A_2，A_1A_3，A_1A_4，A_1A_5，A_1A_6。如果这五条中有一条，比如说A_1A_2是红的，那么A_0A_1，

A_0A_2，A_1A_2都是红的，也就是A_0，A_1，A_2三人讨论的是同一个题目。

如果这五条线段都不是红的，其中必有三条同色。不妨假设A_1A_2、A_1A_3、A_1A_4是黄色。如果A_2A_3、A_3A_4、A_2A_4中有一条为红色，那么它与A_0引出的两条红线段组成一个红色三角形；如果其中有一条为黄色，那么它与A_1引出的两条黄线段组成黄色三角形；如果三条全为蓝色，那么$\triangle A_2A_3A_4$就是蓝色三角形，总之结论成立。

"涂色"已成为符号处理技术中的一种典型方法。这种讨论一个图中是否存在同色三角形（或其它同色图形）的问题在图论中就称为拉姆齐问题。这类问题内容极其丰富。

下面是一个讨论图中是否存在同色矩形的例子：

设n为自然数，不大于44。证明对每个定义在N^2上，值在集$\{1, 2, \cdots, n\}$中的函数f，存在四个有序数对(i, j)，(i, k)，(l, j)，(l, k)，满足

$$f(i, j)=f(i, k)=f(l, j)=f(l, k)$$

其中i，j，l，k是这样的自然数：存在自然数m，p，使

$$1989m\leqslant i<l<1989+1989m$$

$$1989p\leqslant j<k<1989+1989p$$

（第30届国际数学竞赛预选题，古巴提供）

分析　将函数值为$t(1\leqslant t\leqslant n)$的点染上第$t$种颜色。问题即将正方形

$$\{(x, y) | 1989m\leqslant x<1989(m+1),$$
$$1989p\leqslant y<1989(p+1)\}$$

中的整点染上颜色，证明在颜色种数$\leqslant 44$时，必存在一个边与坐标轴平行的矩形，其四个顶点是同一种颜色。

由于正方形中有1989^2个整点，因而至少有$\left[\frac{1989^2}{44}\right]+1=q$个点涂上同一种颜色。所以只需证明将正方形中$q$个点染上红色时，必有一个顶点为红色的矩形，它的边平行于坐标轴。

设第i列中有a_i个点染上红色，则

$$\sum_{i=1}^{1989} a_i=q=\left[\frac{1989^2}{44}\right]+1 \tag{1}$$

在第i列，有$C_{a_i}^2$对点，每一对由两个红点组成。如果

$$\sum_{i=1}^{1989} C_{a_i}^2>C_{1989}^2,$$

那么必有两列，这两列中各有一对红点在相同的两行上，也就是四个点构成一个合乎要求的矩形。

由柯西不等式

$$\begin{aligned}\sum_{i=1}^{1989} C_{a_i}^2&=\frac{1}{2}\sum_{i=1}^{1989}(a_i^2-a_i)=\frac{1}{2}\left(\sum_{i=1}^{1989}a_i^2-q\right)\\&\geqslant\frac{1}{2}\left(\frac{\left(\sum_{i=1}^{1989}a_i\right)^2}{1989}-q\right)=\frac{1}{2}\left(\frac{q^2}{1989}-q\right)\\&=\frac{q}{2\times1989}(q-1989)\\&\geqslant\frac{1989}{2\times44}\left(\frac{1989^2}{44}-1989\right)\\&=\frac{1989^2}{2\times44^2}\times1945\\&>\frac{1989^2}{2}>C_{1989}^2,\end{aligned}$$

因此，结论成立。

三、一类丢番图方程

对不定方程作广泛而深入研究的第一个人是古希腊数学家丢番图。现在，人们把求整数解的整系数不定方程称之为“丢番图方程”。丢番图没有给出不定方程的一般解法，后世很多数学家在这方面作了努力。我们知道，1900年德国大数学家希尔伯特提出了23个著名的数学难题，其中第10个问题就是关于丢番图方程的：“是不是可以设计一种计算步骤，以判定一个整系数方程有没有整数解？”这个问题的答案是否定的。

这里只讨论一类特殊的丢番图方程。例如：

求方程$x+y+z+u=100$的非负整数解的组数。

我们这个例题中的不定方程是一次的，另外，需求的是非负整数解的组数。尽管如此，用通常解方程的办法来研究也是相当麻烦的。我们设法对它进行符号处理。

设$x=a$，$y=b$，$z=c$，$u=d$是方程的一组非负整数解，考虑到a，b，c，d都是非负整数，我们就可写出下面形式的一个等式：

$$\underbrace{1+1+\cdots+1}_{a个}+\underbrace{1+1+\cdots+1}_{b个}+\underbrace{1+1+\cdots+1}_{c个}$$
$$+\underbrace{1+1+\cdots+1}_{d个}=100$$

可见，对于每一组非负整数解$x=a$，$y=b$，$z=c$，$u=d$，我们均可构造如下一串符号

$$\underbrace{1\ 1\cdots 1}_{a个}\ 0\ \underbrace{1\ 1\cdots 1}_{b个}\ 0\ \underbrace{1\ 1\cdots 1}_{c个}\ 0\ \underbrace{1\ 1\cdots 1}_{d个}$$

与之对应，其中$a+b+c+d=100$，反之，这种“一串符号”也对应于一组非负整数解。

这个符号有什么特点？仔细观察、分析，就可发现，这是在103个空位上，任选三个空位写上“0”，而其余空位上均写上“1”。原问题转化为“不尽相异元素的排列问题”。

因此，非负整数解的组数为

$$C_{103}^{3}=\frac{103!}{100!\ 3!}=176851$$

上述方法可用以解决一类计数问题。例如：

高二年级8个班协商组成年级篮球队，共需10名队员，每个班至少要出1名，有多少种不同组成方式？

本题相当于求如下的不定方程

$$x_1+x_2+\cdots+x_8=10$$

的不同的正整数解(x_1，x_2，…，x_8)的组数。

它又等价于求不定方程

$$x_1+x_2+\cdots+x_8=2$$

的不同的非负整数解（x_1-1，x_2-1，…，x_8-1）的组数。

由上例所述的方法可知共有

$$C_9^7=36$$

种不同的组成方式。

当然本题也可有其它的解法。例如将问题区分为：“有一班出3人，其他班各出1人”，“有两个班各出2人，其他班各出1人”等情况来分头计算，然后相加。但相比之下，解答过程要繁琐得多。可见上述解法确实具有其优越性。

下面再看一个排列与组合的综合题：

晚会上共有6个演唱节目和4个舞蹈节目，要求每两个舞蹈节目之间至少有一个演唱节目，共有多少种不同的节目顺序

表?

这问题曾被选为高考试题。我们可先从组合入手，再考虑排列。先进行符号处理，将6个演唱节目视为6个一字排开的小圆



圈，要在它们之中插入4根“小棒”“1”（代表舞蹈节目），“小棒”不能相邻，但可放在所有圆圈的两头，如此，共有7个空位，故可有C_7^4种插法。然后再考虑演唱节目之间的排序、舞蹈节目之间的排序，从而由乘法原理知，共有

$$C_7^4 \cdot P_6^6 \cdot P_4^4 = 604800$$

种不同的节目顺序表。

四、利用函数符号解决问题

几十年前，F·克莱因领导的数学教育改革运动在德国“引起了一场很大的骚动”。这个运动所采用的口号是“用函数来思考”。改革者宣称：一般受教育者在数学课上应该学会的重要事情是用变量和函数来思考。改革者的口号有一定的道理，因为函数贯穿于数学理论和应用的每一个场合。在很多问题中，其元素间存在某种对应关系，发现了这种对应关系，就可把原问题用数学语言表达出来，并借助于函数的表达式或图象符号帮助解决。

例 在n名选手的循环赛中，每两人比赛一次（无平局）。证明以下情况恰有一种发生：

（1）可将选手分为两个非空集合，使得一个集合中的任一名选手战胜另一个集合中的每一名选手。

（2）所有选手可以标上号码1至n，使得第i名选手战胜第$i+1$名，而第n名战胜第1名。

分析 先引用集合与函数的符号，对问题进行符号处理。原问题等价于：

对有限集x的每两个不同元素的有序对（x，y），有一个数$f(x, y)=0$或1与之对应，并且对所有$x, y(x\neq y)$，$f(x, y)\neq f(y, x)$。证明以下两种情况恰有一种出现：

（1）X是两个不相交的非空集合U，V的并，对于任意$u\in U$，$v\in V$均有$f(u, v)=1$。

（2）X的元素可标上x_1，x_2，…，x_n使得

$$f(x_1, x_2)=f(x_2, x_3)=\cdots=f(x_{n-1}, x_n)$$
$$=f(x_n, x_1)=1$$

再用曲线符号来代替函数符号：

在$f(x, y)=1$时，画一条从点x至y的（有向）弧$x\to y$。在$f(x, y)=0$时（这时$f(y, x)=1$），画一条从y至x的弧$y\to x$，产生一个有向图（按原题的说法，$x\to y$即选手x胜选手y）。

(1)，(2)显然不可能同时出现（若(2)出现，则图成为一个圈，那么，从$v\in V$立即得出所有点均属于V，(1)不能出现），只需证(1)，(2)中至少有一种出现。

$n=2$时，(1)显然成立。在$n\geqslant 3$时，设x_1胜得最多，若x_1胜所有选手，则(1)成立（取$U=\{x_1\}$）。否则设y胜x_1，被x_1战胜的选手中必有z胜y（因为x_1胜得最多），这样图中就有圈存在。

设最长的圈为$x_1\to x_2\to\cdots\to x_m\to x_1$（$i\neq j$时，$x_i\neq x_j$）。若$m=n$，(2)成立。我们设$m<n$。

令$C=\{x_1, x_2, \cdots, x_m\}$，对$y\overline{\in}C$，或者每一$i$均有$x_i\to y$，或者每一$i$，均有$y\to x_i$。否则将有某个$i$，使$x_i\to y$而$y\to x_{i+1}$（约定$x_{m+1}=x_1$），从而圈可以扩大为

$x_1\to x_2\to\cdots\to x_i\to y\to x_{i+1}\to\cdots\to x_m\to x_1$出现矛盾。令

$$A=\{y\in X/C: y\to x_i, i=1, 2, \cdots, m\}$$

$$B=\{y\in X/C: x_i\to y, i=1, 2, \cdots, m\}$$

则$A\cup B=X/C$，从而A，B至少有一个非空。

对任一对$a\in A$，$b\in B$。若有$b\to a$，则

$$a\to x_1\to\cdots\to x_m\to b\to a$$

是更大的圈，矛盾。因此恒有$a\to b$，令（若B非空）

$$U=A\cup C,\ V=B,$$

或令（若A非空）

$$U=A,\ V=B\cup C,$$

则(1)成立。

五、利用向量解决问题

向量概念源出于一些物理量。有些量可以用一个数值确定，例如质量、温度、密度、面积和体积等等。这一类量叫数量。还有一些量，只知道它们的数值大小是不够的，要完全表示它们，必须同时说明它们的方向，例如力、速度、加速度等等就是属于这一类。1788年法国数学家拉格朗日的《分析力学》发表。在这本书中，他把这些物理量数学化，即用数学的方法来表示这些量。例如，他用具有确定长度和方向的有向线段来表示一个力f，并沿坐标轴把f分解为三个分力f_x、f_y、f_z。这些分力

作为坐标轴上的有向线段，可以简单地用数来表示。这样，在力学中关于力、速度及加速度的所有方程，可以转变为联系它们的分量的、关于x、y、z的三个方程。拉格朗日没有用“向量”这一名词。

拉格朗日之后，随着电学的发展，在数学和物理学中更广泛地研究了这种有向线段的一般理论。到了19世纪，德国数学家拉格斯曼引入有向线段的记号，并称之为向量。

向量理论由于在力学、物理和技术中的重要性，而构成了解析几何学中的一个重要部分——向量代数。这一名称是为了区别于用分析方法研究向量的理论——向量分析，后者是由美国数学家吉布斯于1881——1884年建立的。

下面我们选择一个初等问题，利用向量来处理问题。

例　A，B，C，D在半径为1的球上。已知$AB\cdot BC\cdot CA\cdot DA\cdot DB\cdot DC=\frac{512}{27}$，证明$ABCD$是正四面体。

分析　设球心为O，连OA，OB，OC，OD，则需证明OA，OB，OC，OD四线段两两夹角相等，然而与夹角概念紧密联系的就是方向问题，这就使我们联想到向量。

记球心O到A，B，C，D的向量为$\vec{e}_i$（$i=1$，2，3，4），则$|\vec{e}_i|=1$，且

$$\prod_{1\le i<j\le 4}\left|\vec{e}_i-\vec{e}_j\right|=\frac{512}{27}$$

即

$$\left(\frac{2^9}{3^3}\right)^2=\prod_{1\le i<j\le 4}(\vec{e}_i-\vec{e}_j)^2$$

$$=\prod_{1\le i<j\le 4}(2-2\vec{e}_i\vec{e}_j)$$

约去2^6得

$$\frac{2^{12}}{3^6}=\prod_{1\leqslant i<j\leqslant 4}(1-\vec{e}_i\vec{e}_j) \tag{1}$$

由(1)及算术—几何平均不等式，得

$$\frac{2^2}{3}\leqslant\frac{1}{6}\sum_{1\leqslant i<j\leqslant 4}(1-\vec{e}_i\vec{e}_j)$$

$$=1-\frac{1}{6}\sum_{1\leqslant i<j\leqslant 4}\vec{e}_i\vec{e}_j$$

即 $$4+2\sum_{1\leqslant i<j\leqslant 4}\vec{e}_i\vec{e}_j\leqslant 0$$

$$\left(\sum_{i=1}^{4}\vec{e}_i\right)^2\leqslant 0$$

$$\sum_{i=1}^{4}\vec{e}_i=0 \tag{2}$$

又由(1)得

$$\frac{2^8}{3^4}=\sqrt[3]{\prod_{1\leqslant i<j\leqslant 4}(1-\vec{e}_i\vec{e}_j)^2}$$

$$=\prod_i\sqrt[3]{\prod_{j\neq i}(1-\vec{e}_i\vec{e}_j)}$$

$$\leqslant\prod_i\left[\frac{1}{3}\sum_{j\neq i}(1-\vec{e}_i\vec{e}_j)\right]$$

$$=\prod_i\left(1-\frac{1}{3}\vec{e}_i\sum_{j\neq i}\vec{e}_j\right) \tag{3}$$

利用(2)式结果，上式就等于

$$\prod_i\left(1+\frac{1}{3}\vec{e}_i\vec{e}_i\right)=\left(\frac{4}{3}\right)^4$$

于是(3)中的$\leqslant$取等号，并且对所有$i\neq j$，

$$1-\vec{e}_i\vec{e}_j=1+\frac{1}{3}$$

即 $$\vec{e}_i\vec{e}_j=-\frac{1}{3} \tag{4}$$

（4）表明OA，OB，OC，OD两两夹角相等，所以这四面体$ABCD$的各棱均相等，它是正四面体。

六、构造几何图形解题

构造法在数学研究中有广泛的应用。古希腊毕达哥拉斯学派的希伯索斯，通过构造边长为1的正方形的对角线，找到了客观存在的无理数。德国著名数学家康托（Cantor）通过构造康托集，证实了存在不可数但测度为零的集合，为数学的发展做出了重大贡献。

作为构造法中的辅助对象可以是多种多样的，其中，引入辅助图形是重要方法之一。

我们知道，几何图形也是符号，因此，构造几何图形实质上也是引入符号。

古代数学家早已懂得构造几何图形解题。众所周知，对于一元二次方程

$$ax^2+bx+c=0 \qquad (a\neq 0)$$

可以先把常数项移到方程的右边，再把左边配成一个完全平方式，如果右边是非负常数，就可以进一步通过直接开平方来求出它的根。这种解二次方程的方法叫做配方法。

配方法源出于古代的图解法。无论是我国，或是希腊、阿拉伯、印度等，最初对二次方程都常常用图解的方法求根。下面我们以830年的阿拉伯数学家阿尔·花拉子模的解法为例，他就是使用这种图解法的代表。

比如解方程

$$x^2+10x=39$$

他的解法如下：

如图5-11，设AB表示未知量x的值。作正方形$ABCD$，延长DA到H，延长DC到F，使$AH=CF=5$，这里5是x的系数10的一半。以DH及DF为边作正方形，于是Ⅰ、Ⅱ、Ⅲ三块面积各是x^2、$5x$及$5x$。这三者之和就是方程的右边(39)。现在把两边加上面积Ⅳ即25。因此整个正方形面积39+25即64，它的边DH必是8。于是AB或AD等于8－5即3，这就是x的值。这种图解法实质上就是现在大家熟悉的配方法。

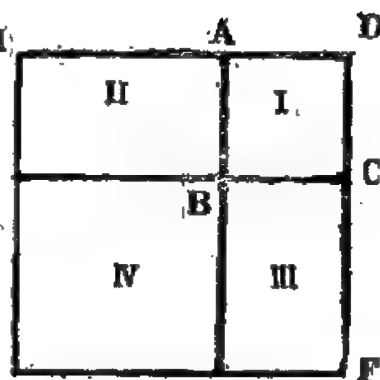


图5-11

$$x^2+10x+25=39+25$$

$$(x+5)^2=64$$

$$x+5=\pm 8$$

故 $$x_1=3,\ x_2=-13$$

显然，图解法的实质就是引入几何符号解决问题，它只能取得正数根。

可以看出，在问题解决过程中，要正确地构造辅助图形解题，除了要熟悉有关几何图形的性质外，还要有一定的逻辑推理与直觉洞察能力。

例 设a、b、c、$d\in R^+$，$\frac{a}{b}=\frac{c}{d}$，且a为最大，求证$a+d>b+c$。

由$ad=bc$，可联想到圆的割线。考虑到a为最大，设法构造一个圆，将a作为通过直经的割线，在其中寻找$a+d$和$b+c$，

比较其长短。这是可以做到的。

如图5-12，取直线ABC，使$AC=a$，$AB=d$。以BC为直径作半圆O，不妨设$b\geqslant c$，作割线$AD=b$，交圆于E，过O作弦ED的垂线OF，F为垂足。

$$\because \quad AC\cdot AB=AD\cdot AE$$

$$\therefore \quad AE=c$$

又 $$AO=AB+\frac{BC}{2}=d+\frac{a-d}{2}=\frac{a+d}{2}$$

$$AF=AE+\frac{ED}{2}=c+\frac{b-c}{2}=\frac{b+c}{2}$$

在$\triangle AOF$中，$AO>AF$

$\therefore \quad a+d>b+c$

有时，构造辅助图形还需要更复杂的构思过程。

例 若$0<\theta<\frac{\pi}{2}$，求证$\sin\theta+\cos\theta\leqslant\sqrt{2}$。

原式可变形为

$$\frac{\sqrt{2}}{2}\cdot\sin\theta+\frac{\sqrt{2}}{2}\cos\theta\leqslant 1$$

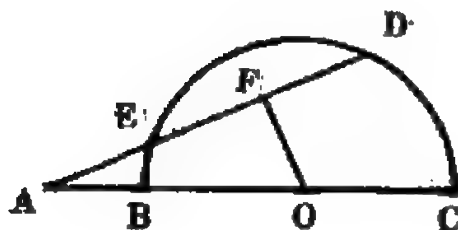


图 5-12

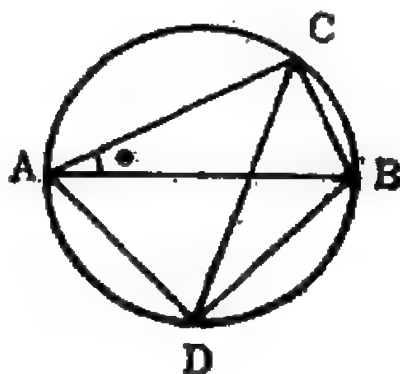


图 5-13

由这个不等式的结构形式可联想到托勒密定理。下面我们设法构造圆内接四边形，因为式中 1 为最大，故取圆直径为 1。

如图5-13所示，作直径为1的圆，设AB为直径，作$\angle CAB=\theta$（可以是0到$\frac{\pi}{2}$的任意角），$\angle ABD=\frac{\pi}{4}$，C、D为圆上的点。则

$$AC=\cos\theta,\ BC=\sin\theta$$

$$AD=BD=\frac{\sqrt{2}}{2}$$

由托勒密定理，得

$$BC\cdot AD+AC\cdot BD=AB\cdot CD$$

即

$$\frac{\sqrt{2}}{2}\sin\theta+\frac{\sqrt{2}}{2}\cos\theta=CD\leqslant 1$$

因此，$\sin\theta+\cos\theta\leqslant\sqrt{2}$

上述解法在构思上确实是十分巧妙的，固定了直径AB，有了直角三角形，正弦和余弦值转化为圆内接四边形的边。如此，非常成功地应用了托勒密定理。

* * *

“问题解决”中的符号处理技巧，举不胜举，不可能从中概括出一个统一的模式，但总的讲来，都离不开对问题进行模式化抽象。每个问题总要涉及一些对象及对象之间的关系，只有通过观察和分析，找到它们的本质属性和特征，舍弃其它非本质的内容，然后用适当的符号来表示，才能抓住问题的实质，获得合适的数学模型。

第六章　注目于现代数学教育

——重视符号教学

"如果一个学生要成为完全合格的、多方面武装的科学家，他在其发展初期就必定来到一座大门并且必须通过这座门。在这座大门上用每一种人类语言刻着同样的一句话：'这里使用数学语言'。"这段话极其形象地描绘了掌握数学语言的重要性。

斯托尼亚尔十分重视数学语言的教学。他认为，数学教学也就是数学语言的教学，他写道：

"如果我们同意，数学在某方面是描述其他科学和实践活动中产生的实际情况的专门语言，同意解决数学以外产生的问题首先要把这些问题翻译成数学语言并且把所得结果再从数学语言翻回原来那个学科领域的语言，最后，如果认为懂得数学就意味着会用它去解决生活中、各科学技术领域中以及实践活动中产生的各种问题，那么，十分清楚，数学教学也就是数学语言的教学。"

搞好数学语言的教学的关键是搞好数学符号的教学，科学发展的形势要求数学教育必须重视数学符号的教学。

§1　符号学习的心理分析

调查表明，人们在学习数学符号的过程中或多或少地总存

在一些心理障碍。为了掌握数学符号，必须排除心理障碍。学习数学符号的心理障碍，有其产生的客观背景，也有其主观因素。客观背景是源于数学符号的抽象性；主观因素是来自思维定势的影响。

定势也叫“心向”，是先于某种活动而指向某种活动的动力准备状态。例如赛跑时的起跑姿势，便是一种定势。在学习过程中，个体应用知识的准备状态，也是一种定势，这是思维定势。思维定势往往表现为用固定的思路和习惯去考虑问题与解决问题。例如，学生对因式分解“$a^{n+1}-3a^{n}+2a^{n-1}$”感到困难，这是因为他们习惯于把$a^{3}\cdot a^{2}$化成a^{5}；而不善于将a^{5}化成$a^{3}\cdot a^{2}$的缘故。

再如，化简“$\sin(x-y)\cos y+\cos(x-y)\sin y$”时，有些学生总习惯于先将$\sin(x-y)$与$\cos(x-y)$分别展开后再进行计算，而不善于将$(x-y)$与$y$看成两个单角进行计算。

思维的定势，有时起积极作用，有时则起消极作用。

学习数学符号的心理障碍主要表现为下面几种：

一、情绪障碍

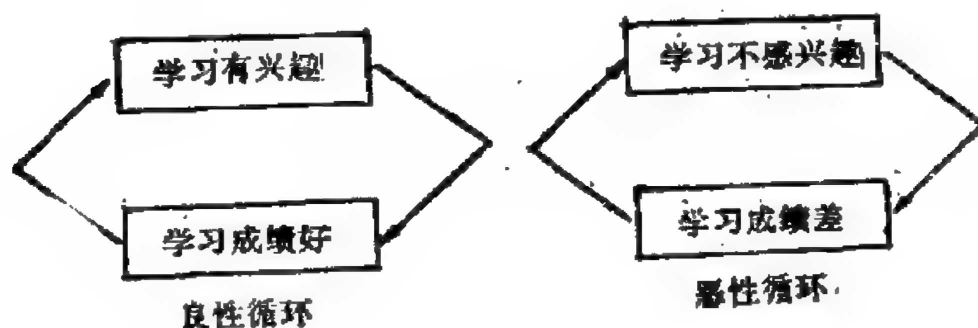
所谓情绪，就是个体受到外部环境的刺激而产生的一种心理状态或心理反映。情绪的产生是以客观事物和对象是否满足个体的需要为中介的。通常那些满足个体需要的对象，会引起满意、高兴、喜悦等积极情绪；反之妨碍需要得到满足的对象，就会引起痛苦、忧愁、厌恶等消极情绪。

在学习过程中，学习情绪，直接影响学习兴趣。

兴趣是学习活动中一个重要的心理因素，是人们爱好某种活动或力求认识某种事物的倾向。因此，兴趣是带有情绪色彩

的意向活动。巴甫洛夫把兴趣视为增强紧张度、引起大脑皮层活动状态的因素。符合兴趣的工作，容易取得成果。美国心理学家布鲁纳说：“最好的学习动因是学员对学习材料有内在兴趣。”孔子说：“知之者，不如好之者，好之者，不如乐知者。”爱因斯坦说：“兴趣是最好的老师”。

人们多次调查表明，学习兴趣和学习成绩之间存在着一种“良性循环”和“恶性循环”：



良性循环　　恶性循环

古今中外许多科学家的成才都是始于兴趣，得益于良好的情绪。

学生对于数学符号的情感，直接影响数学符号的学习效果。数学家A·巴特斯布（A·Battersby）评价他们国家的学生说：“实际上，我们的学校的成绩在一个方面常常是消极的，那就是学生们学习后不但对数学符号冷漠，而且感到它们可怕。”这种现象也许是带有一般性的。这种情绪障碍主要来自两个方面。

第一，因为情绪的产生是以客观事物和对象是否满足个体的需要为中介的。数学符号的高度抽象性，使学习者不能立即

感到“满足个体的需要”，相反地，往往还会因其抽象、难懂而产生沮丧心情。对于许多非数学专业的人来说，“学习数学语言可能要比学习任何一种外语都困难得多。”对于看不懂数学符号的人来说，古怪、离奇的数学符号就象“天书”一样令人望而生畏。正如数学家H·脱普粒茨（H·Toeplitz)所说：“数学由于其语言、标记方法和奇怪的符号，像是被一座高墙与外世隔绝。对于墙外的人，他觉得墙里面大部分都是秘密，是沉闷、平凡的数字，是那些按照不可逃避的必然定律起作用的、没有生气的数学家。”那些“墙外的人”，对于数学符号当然不会产生喜悦的积极情绪。

第二，一些不适当的、夸大了的宣传，歪曲了数学符号的形象，使学习者产生一种思维定势。

数学家H·维尔（H·Weyl）曾说过：“数学因它总是以抽象的方式来讨论问题而弄得声名狼藉。其实这个坏名声只有一半是该当的。”数学符号是抽象的，但它充满生机，有其数学思想，不是枯燥的。然而“公众的舆论”有时并不是“公正”的。有些好心的教师告诫初一的学生说：“数学抽象、枯燥，你们要好好学习，否则将会留级。”这种讲法没有积极作用，只能使学生讨厌数学。正如波里亚所说：“…数学在各个课程中是最不得人心的一门功课，其名声不佳，…。未来的教师通过初级学校学会讨厌数学….他们返回初级学校又教育新的一代讨厌数学。”

二、学习的负迁移

迁移是一种心理现象，是一种学习对另一种学习所产生的影响。

学习能够迁移。早在两千多年前，孔子就提出：“举一隅，不以三隅反，则不复也”，“回也，闻一以知十。”意思是说学习可以“举一反三”、“触类旁通”、“由此及彼”地迁移。19世纪末，20世纪初，有人借助于实验对迁移进行了研究。到了本世纪60年代，美国心理学家布鲁纳第一次把迁移问题作为教育问题的核心提到日程上来。之后，受到各国的心理学家和教育家的关注，甚至把它作为一个教育、教学的原则，提出要“为迁移而教”。

学习之间的影响有时是积极的，有时是消极的。凡一种学习对另一种学习起促进作用的，叫做正迁移。凡一种学习对另一种学习起干扰或抑制作用的，称为负迁移。

思维的定势可以促进正迁移的发生，也可能促进负迁移的发生。这主要取决于定势与所要解决的问题是否互相适应。如果定势与所要解决的问题相适应，则产生正迁移；否则，就产生负迁移。

在数学学习中，产生正迁移的现象是经常发生的。例如，学习方程的知识有助于学习不等式；学习数的运算规则有助于学习“式”的运算规则；学会了解一元一次方程，就有助于学习解一元二次方程。

在数学学习中也会产生负迁移。例如，将“方程两边同乘以一个不等于零的数或式，方程的解不变”这一原理，套用到解不等式问题中，作为解不等式的原理，就产生了负迁移。

表现在数学符号的学习方面，有时由于数学符号的形式结构与语义内容脱节，从而也会产生负迁移的现象。例如，乘法对于加法的分配律公式

$$a(b+c)=ab+ac$$

是学生所熟知的，如果忽视了$\sin(\alpha+\beta)$，$\lg(a+b)$的语义内容，就会套用分配律的公式，得出

$$\sin(\alpha+\beta)=\sin\alpha+\sin\beta$$

$$\lg(a+b)=\lg a+\lg b$$

的错误。

同样，诸如：

$$(a\pm b)^2=a^2\pm b^2$$

$$\sqrt{a^2+b^2}=a\pm b$$

$$(U\cdot V)'=U'V'$$

$$\left(\frac{U}{V}\right)'=\frac{U'}{V'}$$

等错误也是源出于学习的负迁移。

*　　*　　*　　*

现代的数学教学必须采取有效的方法排除上述种种心理障碍。正如托普粒斯(O·Toeplitz)所说：“我们只要能消除许多人从童年经历中得到的那种对数学的反感，就能激发这些人对数学的兴趣。”

§2　概念教学的组成部分

恩格斯曾说过：“在一定意义上，科学的内容就是概念的体系。”数学是由概念与命题等内容组成的知识体系。正确理解数学概念是掌握数学基础知识、获得数学能力的前提，学生的逻辑思维能力、空间想象能力、运算作图能力以及“问题解决”过程中的探索能力等等无一不是以清晰、确定的概念为基础的。从一定意义上来说，数学水平的高低，取决于对数学概念的掌

握程度。

符号对于构成数学概念起着极其重要的作用。

有些数学概念的构成是先用自然语言表达，随后赋予符号表示。例如函数概念等。

有些抽象层次较高的数学概念的形成，必须借助于符号给出定义。例如

绝对值的概念是通过一系列数学符号表示的实例：$|2|=2$；$|-2|=-(-2)$；$|0|=0$，概括出绝对值的属性，定义为

$$|a|=\begin{cases} a & a>0 \\ 0 & a=0 \\ -a & a<0 \end{cases}$$

导数的概念是对一系列数学符号表示的实例，如

瞬时速度 $\lim\limits_{\Delta t\to 0}\dfrac{\Delta s}{\Delta t}=\lim\limits_{\Delta t\to 0}\dfrac{f(t+\Delta t)-f(t)}{\Delta t}$；

局部密度 $\lim\limits_{\Delta x\to 0}\dfrac{\Delta m}{\Delta x}=\lim\limits_{\Delta x\to 0}\dfrac{m(x+\Delta x)-m(x)}{\Delta x}$；

切线斜率 $\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}=\lim\limits_{\Delta x\to 0}\dfrac{f(x+\Delta x)-f(x)}{\Delta x}$；

……

进行抽象，从而得到其定义，即

$$f'(x)=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

还有一些数学概念，纯粹是用符号或用由符号组成的表达式来定义的。例如

行列式与矩阵概念；

幂函数、指数函数、对数函数、三角函数、反三角函数；

某些用含参量积分定义的函数，如Γ函数、β函数、ψ函

数、椭圆函数；

各种积分变换，如拉氏变换、傅氏变换、勒让德变换；

……

既然，数学符号与数学概念有如此不可分割的关系，数学符号的教学就理该是概念教学的重要组成部分。然而，这问题尚未引起人们的普遍重视。

爱因斯坦曾说过："一个人的智力发展和他形成的概念的方法，在很大程度上是取决于语言的。"可见语言和概念、语言和智力的密切关系。数学的语言是由符号组成的语言，在概念教学中，必须重视符号的教学。

一、重视语义分析

在概念教学中，必须重视对符号的语义分析。符号只是代表概念的物质外壳，如果学生不了解符号的涵义，那就什么也不知道。而且对于一个符号，如果学生只是一知半解的使用它，则也是无法掌握和运用自如的。正如斯托尼亚尔所说，如果学生不理解数学语言表达式的意义，就"不能把非数学问题化成数学问题，他们的知识将是形式主义的、无益的"。在教学过程中，要自始至终给表示概念的符号赋于具体内容。

有些数学表达式之间，在形式结构上似乎没有多大差别，但其语义内容却完全不一样。例如：

$$a_1+a_2+\cdots+a_n$$

与

$$a_1+a_2+\cdots+a_n+\cdots$$

似乎相差无几，但两者却代表完全不同的数学概念。前者代表代数和，是有限运算；后者已经不是代数和，它被定义为"部

分和"的极限，属于无限运算。这可算是"差之毫厘，失之千里"。

教师必须帮助学生透过符号、表达式的形式结构，了解其本质内容。这是帮助学生掌握概念的需要也是培养运算能力的需要。例如：

已知：$$H(x)=\begin{cases}0 & 当x\leqslant 0\\ 1 & 当x>0\end{cases}$$

画出函数$y=H(x-1)$的图象。

这是1984年全国高考理工科试题。有些考生竟束手无策，其原因之一就是对于函数符号

$$y=f(x)$$

的含义和实质理解不深。

再如，初学对数的人，往往对公式

$$a^{\log_a N}=N \qquad (a>0,\ a\neq 1,\ N>0)$$

不甚理解，其实，只要真正理解对数符号$\log_a N$的含义，就能直接洞察到该公式的正确性。

苏联数学家鲁金在谈到欧拉时说："欧拉的洞察力是那样深邃，不论多么复杂深奥的公式，在他强有力的手里都变得服服贴贴，宛如柔软蜂蜡一样；在他的威力面前都得规规矩矩地献出一切。……他可以本能地直接感觉到公式里的真理与虚假，他调迁公式的技巧，对公式进行数量上的评估与变换的功夫，对结果的本质瞬间猜中的本领——这些都令人叹为观止。可以毫不夸张地说，在欧拉眼里，数学公式本身自始至终都充满了生命力，讲述着有关自然现象的最深刻的东西。只要他一碰到公式，就能使公式由'哑吧'变成会说话的人，并能作出饱含深

邃含义的回答。”当然，像欧拉这样的数学大师，毕竟是少数。但是，现代的数学教育在培养“通才”的同时也应该努力培养未来的符号大师。

我们知道，通过操作数学符号，便能使某些数学操作过程达到符号化、形式化、自动化；思维过程出现约简、越层现象。但是这种操作毕竟不同于机械操作，这种操作在自动工作的同时，操作者应对符号代表的概念属性了解透彻，不允许有丝毫的含糊和混淆。

斯托尼亚尔曾指出，传统数学中学生知识表面化的根源往往是，数学语言的学习中形式和内容的脱节，其实质就是数学语言的符号和公式与它们所表示的东西脱节。这也就是在教学中对数学语言的语义注意不够的结果。如此，学生在从具体材料抽象概括出抽象结论时，容易受具体对象的非本质特征的干扰而不能正确理解抽象结论的语义内容，例如：

从$\sqrt{2}$不是有理数引进无理数的概括，学生往往就把无理数理解为不尽方根数；

由$|2|=2$，$|-2|=2$，$|0|=0$等具体材料概括出绝对值的概念，学生往往认为$|-a|=a$。

学生如果只知道符号形式而不理解其涵义，就很容易造成学习的负迁移，符号将阻障学生的思维发展和对数学知识的掌握。例如，如果不理解函数符号$y=f(x)$中各字母的涵义时，类似于$f(x_1+x_2)=f(x_1)+f(x_2)$的错误就可能发生。

如果不理解无穷级数的和的涵义及其性质，而把有限运算的性质套用到无限运算之中，也会产生负迁移。例如：

级数

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}-\frac{1}{8}+\cdots$$

是收敛的，设其和为l。因而

$$2l=\frac{2}{1}-\frac{1}{1}+\frac{2}{3}-\frac{1}{2}+\frac{2}{5}-\frac{1}{3}+\frac{2}{7}-\frac{1}{4}+\cdots$$

我们知道，对于任意有限多项的和而言，可以任意地变换其项的次序而不改变其和的值。现在，如果适当变换诸项的次序，以使级数中分母相同者得以合并，于是应有

$$\begin{aligned}2l=&\frac{2}{1}-\frac{1}{2}+\frac{2}{3}-\frac{1}{4}+\frac{2}{5}-\cdots\\&-\frac{1}{1}\qquad-\frac{1}{3}\qquad-\frac{1}{5}-\cdots\\&=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\cdots\end{aligned}$$

如此，得到$2l=l$。但从原级数的头几项就可看出$l\neq0$，于是$1=2$。这显然是一个荒谬的结论。

类似性质的错误还表现于求极限运算之中，例如：

$$\begin{aligned}&\lim_{n\to\infty}\frac{1+2+\cdots+n}{n}=\lim_{n\to\infty}\left(\frac{1}{n}+\frac{2}{n}+\cdots+\frac{n}{n}\right)\\&=\lim_{n\to\infty}\left(\frac{1}{n}+\frac{2}{n}+\cdots+\frac{n}{n}\right)\\&=0+0+\cdots+0=0\end{aligned}$$

错误在于对所求极限式的涵义理解不深。事实上，这里的和式$\frac{1}{n^2}+\frac{2}{n^2}+\cdots+\frac{n}{n^2}$的项数不确定，随$n$的增大而无限增多，因此，不能直接用数列极限的加法法则来计算。

“语义分析”包括分析数学符号的内在条件，数学符号的出现往往是伴随着一定的条件的。例如：

a^0存在的条件是$a\neq0$；

$\log_a N$存在的条件是$a>0$且$a\neq1$，$N>0$；

$y=\arcsin x$存在的条件是$|x|\leqslant 1$，$|y|\leqslant\frac{\pi}{2}$，

等等。

忽视条件，就会造成形式与内容的脱节，就会造成错误。

数学表达式的涵义往往与其中某些字母的取值大小密切相关。例如，在极坐标系中圆锥曲线的方程可表为

$$\rho=\frac{ep}{1-e\cos\theta}$$

该表达式的具体涵义，随着其中的字母e——离心率的变化而变化：

当$0\leqslant e<1$时，该式表示椭圆；

当$e>1$时，该式表示双曲线；

当$e=1$时，该式表示抛物线。

因此，对字母的各种取值情况进行分析，也是很重要的。

二、注意剖析结构

关于数学符号，除了其抽象性之外，符号结构本身也是造成学生学习困难的原因之一。有些数学概念是用构造法引进的，且构造过程繁琐，其相应的数学符号结构复杂、层次多。导数与定积分就是很典型的例子，两者都是用构造法引进的。所谓导数，就是指

$$\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$$

所谓定积分，就是指

$$\lim_{\|\Delta x_i\|\to 0}\sum_{i=1}^{n}f(\xi_i)\Delta x_i$$

这两个结构式都是特殊类型的极限，的确是结构复杂、层次多。初次接触这些符号的人，往往感到抽象难懂。因此，在概念教学

中，必须注意剖析数学符号的结构。可以从以下几方面进行剖析：

第一，剖析结构的实际来源。数学中的一切符号结构式，都有其产生的实际背景（包括理论背景）及历史背景，教师必须追本溯源，剖析其来源。以定积分为例，它起源于求图形的面积和一些其它量的求和问题。求图形的面积，很早就为数学家们所注意。古希腊的数学家阿基米德早就用分割的方法计算过抛物弓形的面积、球和球冠的面积、螺线下面积等。到了16世纪以后，面积计算的问题又从力学获得新的动力和启发，产生了新的成果，而且在原子论的启示下，普遍把一块任意形状的面积近似地看成很多细窄的矩形面积的总和。后来，人们进一步发现。要求不规则图形的面积，只要求下列形式的极限

$$\lim_{\max|\Delta x_i|\to 0}\sum_{i=1}^{n} f(\xi_i)\Delta x_i$$

还有很多问题，如求变速直线运动的路程、变力所作功、液体的压力等都归结到同样类型的和的极限。因此，人们就必须对这种结构式进行深入研究，从而得到定积分的概念。可见，定积分概念中的符号结构式不是数学家头脑里凭空制造出来的，完全是由于科学与实践的需要，从实际问题的解决过程中总结出来的；是经过人们几百年、几千年的努力，才逐步形成的。

结合历史讲清实际来源，学生就不致感到抽象、枯燥，有利于培养学习情感。

第二，剖析结构的层次。数学中的很多符号结构是具有层次性的，为了使学生彻底掌握其结构，教师必须认真剖析结构式中的几个层次。例如，导数与定积分概念的构造过程中都有三个层次。

导数概念中的三个层次：

1.自由变量的改变量Δx，求函数的改变量

$$\Delta y=f(x_0+\Delta x)-f(x_0)$$

2.求两个改变量的比：

$$\frac{\Delta y}{\Delta x}=\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$$

3.令$\Delta x\to 0$，求极限：

$$\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$$

定积分概念中的三个层次：

1.分割积分区间，作乘积$f(\xi_i)\Delta x_i$；

2.作和式$\sum_{i=1}^{n}f(\xi_i)\Delta x_i$；

3.令$\|\Delta x_i\|\to 0$，求极限

$$\lim_{\|\Delta x_i\|\to 0}\sum_{i=1}^{n}f(\xi_i)\Delta x_i$$

符号结构的层次，反映了数学概念的形成过程，因此，剖析结构的层次，理该是概念教学的重要内容。另外，教师还必须结合实例讲清各个层次的含义。例如，让学生按定义中的三个层次求一些简单函数的导数和定积分。这是引导学生理解符号结构的重要环节。

第三，对结构进行几何解释。希尔伯特高度重视几何解释作用，他指出几何图形能帮助记忆，能"令人想到曾经是形成新概念的缘由的那种现象"。对于很多数学概念的符号结构式，可以赋于几何意义，利用几何直观揭示各层次的意义及相互关系。例如，导数与定积分概念中各个层次就有鲜明的几何意义。在定积分概念中：

$f(\xi_i)\Delta x_i$代表小矩形面积，它是细窄的曲边梯形面积的

近似值；

$\sum_{i=1}^{n} f(\xi_i)\Delta x_i$代表诸小矩形面积的和，它是曲边梯形面积的近似值；

$\lim_{|\Delta x_i| \to 0} \sum_{i=1}^{n} f(\xi_i)\Delta x_i$ 代表曲边梯形的面积。

如果能运用动画片来显示从小矩形面积的和转化到曲边梯形面积的变化、运动过程，那就更加形象化。这样，对提高学生的思维能力无疑是有益的。

第四，从方法论角度分析。我们已多次看到，数学大师们总是在解决问题的过程中，随着思想方法的发展不断更新数学符号的。数学符号的发展与数学方法论的发展两者之间有密切的联系，对数学符号结构式的形成过程，作方法论方面的剖析是必要的。仍以定积分为例：

建立定积分概念是以极限法为基础的。所谓极限法，就是用极限概念分析问题和解决问题的一种数学方法。极限法的一般步骤可概括为：欲考察一个未知量，先设法选择、构造一个与它有关的变量，确认这变量通过无穷过程的结果就是所求的未知量；最后借助于极限运算来得到所求的结果。我国古代数学家刘徽推算圆面积时就是用了极限法的思想，他主要是观察内接正 $3 \cdot 2^n$边形的面积的变化趋势。确认这些多边形的面积的变化结果就是所求圆面积。就像坐标法是解析几何的基本方法一样，极限法是微积分的基本方法。借助极限法，人们可以“从有限认识无限”，“从近似认识准确”，“从量变认识质变”。

定积分概念的构造过程，体现了微积分的基本思想。例如，怎样求曲边梯形的面积？主要困难是有条边是曲边。如果这条“曲边”换成直线就好办了。因此，我们先“化整为零”：把

曲边梯形的底边分成许多小段，从而曲边梯形也分成许多小窄条；再“以不变代变”或说“以直代曲”：在每个小区间上把函数值近似看作是不变的，相等的，即把小窄条近似看作矩形，求出小窄条的近似面积；再“积零为整”：把各小矩形面积加起来，得到曲边梯形面积的近似值。再“取极限”：分割越细，这些近似值越接近于曲边梯形的面积，通过无限运算，“取极限”使“近似”转化为“准确”，从而得到曲边梯形的面积。这里包含着“化整为零”、“以不变代变”、“积零为整”和“求极限”等过程。这种解决问题的思想方法具有普遍的意义。

在剖析结构层次的同时，注意从方法论角度进行分析。这样做，对于提高学生分析问题、解决问题的能力是有益的。

三、注意名词的剖析

我们知道，人们对数学对象，通过分析、比较，抽象出一类对象的本质属性而形成概念后，总要赋于简便的符号，并用词加以命名即定名称。

概念及其名称之间有本质的区别，它们分属两个不同的范畴。概念是人们认识事物的结果，是属于思想范畴的东西，它反映名称的思想内容。而概念的名称是表达概念的语言形式，是既看得到又听得到的口头语或书写出来的词语或符号。

然而，概念与其名称之间又有不可分割的联系。概念的名称一旦给以确定，它就成为概念的代名词，可使思维过程简缩。借助于概念的名称，人们可以在低级概念的基础上引进高级概念；借助概念的名称，可以对研究对象进行分类。名称，可唤起相应的符号、概念。例如，当我们看到或听到“平行四边形”这个名称，我们就会想到符号“▱”及其概念；当我们看到或

听到“定积分”这个名称，我们就会想到符号

$$\int_a^b f(x)dx$$

及其概念。在人们的思维过程中，概念的名称、符号与概念如此紧紧相连，给人以“三位一体”的感觉，难怪人们常常把名称、符号误认为概念本身。

数学中有些名词用以称呼特定的符号形式，反映了有关概念的特性，教师讲清它们的意义，有利于学生掌握一般规律，更好地掌握符号，理解概念。例如

（1）“一般形式”

在数学中对于式、方程、函数等概念，多是先总结出一般的符号形式，再进行讨论。如一次函数的一般形式、指数函数的一般形式，对数函数的一般形式，一元一次方程的一般形式，一元二次方程的一般形式……。为什么要把某些表达式规定为“一般形式”？对一般形式进行讨论，得到一般的结论，就可用来帮助解决各种具体问题。就一元二次方程来说吧。对其一般形式进行讨论，得到的求根公式、判别式与韦达定理，对于一元二次方程来说都是极其重要的。

（2）最简形式

如最简多项式、最简分式、最简根式等，为什么要把某些符号形式规定为最简形式？事情很简单，人们对于所研究的对象，为了突出其本质的属性，总要尽量在外形上化简。以多项式为例，“合并同类项后的多项式叫最简多项式。”没有最简多项式这个概念，关于多项式的很多问题就难以研究。例如“如果两个最简多项式恒等，那么它们的对应项系数相等。”

这定理说明了两个恒等多项式在外形上是完全一样的，这定理是待定系数法的依据。这里“最简”的条件是必不可缺的。没有最简的条件，本质上完全相同（即恒等）的多项式，在外形上可以千变万化，那么，讨论问题时，就不方便。另外，不规定“最简形式”，在计算问题时，不知道应该化简到那一步为止，就没有统一标准。

（3）标准形式

如抛物线的标准方程，椭圆的标准方程，双曲线的标准方程，自然数的标准分解等。为什么要把某些符号形式规定为标准形式？以椭圆为例，建立不同的坐标系，就可得到不同的方程。为此，同一个椭圆，就会有多种不同的方程。其中，若不规定一个标准的，那么人们就没有共同的语言。因此前人规定在某种坐标系中的椭圆方程称为椭圆的标准方程。

（4）“基本”性质、“基本”定理、“基本”公式

如分式的基本性质、根式的基本性质、积分学的基本公式、代数基本定理等。为什么把这些性质、公式等等加上“基本”二字？顾名思义，所谓“基本”，就是说这些性质、公式是其它许多性质、公式的基础。如分式的基本性质：“分式的分子、分母都乘以或除以同一个不等于零的代数式，分式的值不变。”这条性质是分式进行变形的依据。分式的约分、通分和分式的运算都离不开这条性质。因此，前人把这条性质规定为“基本”性质是非常确当的。关于积分学基本公式，它所起的基本作用，前面我们已经论及。事实上，发现这一公式是积分发展史的一个飞跃。从此，积分学才建立成为一门独立学科。因此，人们把这一公式定名为积分学的基本公式。

关于一些符号名称的变化，也要注意剖析。

对于同一字母符号，在不同的等式中会出现不同的“名称”，对于这些名称的改变，有些学生会感到难以记忆。教师必须讲清名称改变的“实质”，帮助学生理解概念。例如

指数　　　根指数　　　真数

$2^{3}=8$　　$\sqrt[3]{8}=2$　　$\log_{2}8=3$

底数　幂　　被开方数　根　　底数　对数

从以上三式可见，同一数字，在不同等式中，“称谓”也不同。为什么要改变其名称？举例来说，“教师”与“学生”在运动会上就分别改称为“裁判员”与“运动员”，因为他们的地位与作用改变了，职称也相应地改变。数学也如此，由于问题的性质变化了，字母的地位与作用也随着改变。为了使“名称”能更好地反映问题本质，就必须把名称也作相应的改变。

数学中有些名词的来源是很有趣的。例如，“幂”这个词是怎样引进数学的？

众所周知，求n个相同数的积的运算，叫做乘方，乘方的结果叫做幂。

幂的概念的形成是相当曲折和缓慢的。

我国古代幂字至少有十种不同的写法。最简单是“冖”。幂作为名词用是用来覆盖食物的巾，作动词用就是用巾来覆盖。《说文解字》解释说：“冖，覆也，从一下垂也。”

用一块方形的布盖东西，四角垂下来，就成“冖”的形状。将这意义加以引申，凡是方形的东西也可以叫做幂。再进一步推广，矩形面积或两数的积（特别是一个数自乘的结果）也叫做幂。这种推广是从刘徽开始的。

刘徽在263年为《九章算术》作注，在《方田》章求矩形

面积法则下面写道："此积谓田幂"。他还说，长和宽相乘的积叫幂。这是在数学文献中第一次出现幂字。在《勾股》章中，刘徽表述勾股定理为："勾股幂合以成弦幂。"这里幂是指边自乘的结果或正方形的面积。

300多年以后，李淳风重注《九章算术》，还不同意刘徽这样使用幂字。到了明朝，有些数学书中完全不使用幂字。

1607年，利玛窦和徐光启合译欧几里得《几何原本》，在译本中徐光启重新使用了幂字。他说："自乘之数曰幂"。这是第一次给幂这个概念下定义。

另一方面，幂的概念的形成还受到国外的影响。1591年法国数学家韦达的代数名著《分析方法入门》中曾经用拉丁文表达"幂"，以后译成英文后变成意义相当的"*Power*"，1935年，我国出版《数学名词》，把"*Power*"译成"幂"，这个术语从此才算确定下来。

向学生适当介绍名词的来源，能帮助学生建立正确的数学观念，提高学习兴趣。

§3 数学语言的现代化

现代数学涉及的每个研究对象实质上是一个集合，集合论已构成全部数学的基础。集合语言是数学本身的语言，所谓数学语言的现代化就是广泛使用集合语言和数理逻辑语言。不少人反对在中、小学课本中引进集合论和数理逻辑的初步概念及符号。但是，心理学、教育学的研究表明尽早地引入逻辑初步知识是明智的。

一、逻辑运算有助于数学运算

以罗素为代表的逻辑主义学派认为全部数学可以化归为逻辑，因而是逻辑的一个分支；逻辑是符号体系。罗素和怀德海著的三大卷《数学原理》，力图由逻辑规则和逻辑概念出发，演译出全部数学的基本概念和原理。虽然他们并没有完全实现他们的宗旨，但是他们的工作对数学与逻辑的发展作出了重要贡献。他们相当成功地把古典数学纳入了一个统一的合理系统，虽然这个系统不是纯逻辑的，但这样一项工作却成为公理化方法在近代发展中的一个重要起点。

逻辑运算（否定、合取、析取和蕴含）可以用到命题形式上，因此，使用逻辑和集合论的工具有助于数学运算。仅以方程和不等式为例，使用逻辑和集合论的工具有助于理解将复杂的方程和不等式化简的原理。例如：

（1）从方程论的观点看，方程的等价关系可归结为逻辑的等价关系。在解方程和不等式的过程中正是使用了这一条。比如，解方程

$$(x-1)(x-2)=0$$

时，我们确立等价性

$$(x-1)(x-2)=0 \Longleftrightarrow (x-1=0) \vee (x-2=0)$$

并且把析取的真值域定义作组成它的更简单命题形式（线性方程）的真值域的并：

$$\begin{aligned} & \underset{x}{M}\{(x-1)(x-2)=0\} \\ = & \underset{x}{M}\{(x-1=0) \vee (x-2=0)\} \\ = & \underset{x}{M}\{x-1=0\} \cup \underset{x}{M}\{x-2=0\} \end{aligned}$$

$=\{1\}\cup\{2\}=\{1,2\}$

一般地，解方程$f_1(x)\cdot f_2(x)=0$时，我们确立等价性

$$f_1(x)\cdot f_2(x)=0\Leftrightarrow f_1(x)=0\vee f_2(x)=0$$

（在函数f_1和f_2的定义域的交上）并且在此基础上得

$$\begin{aligned}&\underset{x}{M}[f_1(x)\cdot f_2(x)=0]\\&=\underset{x}{M}[f_1(x)=0\vee f_2(x)=0]\\&=\underset{x}{M}[f_1(x)=0]\cup\underset{x}{M}[f_2(x)=0]\\&=E_1\cup E_2\end{aligned}$$

其中 $E_1=\underset{x}{M}[f_1(x)=0]$，$E_2=\underset{x}{M}[f_2(x)=0]$。

(2) 解方程$|f_1(x)|+|f_2(x)|=0$时，我们用等价性

$$|f_1(x)|+|f_2(x)|=0\Leftrightarrow f_1(x)=0\wedge f_2(x)=0$$

所以给定方程的解集通过集合E_1和E_2表示如下：

$$\begin{aligned}&\underset{x}{M}[|f_1(x)|+|f_2(x)|=0]\\&=\underset{x}{M}[f_1(x)=0\wedge f_2(x)=0]\\&=\underset{x}{M}[f_1(x)=0]\cap\underset{x}{M}[f_2(x)=0]\\&=E_1\cap E_2\end{aligned}$$

(3) 解方程 $\dfrac{f_1(x)}{f_2(x)}=0$ 时，我们确立等价性

$$\frac{f_1(x)}{f_2(x)}=0\Leftrightarrow f_1(x)=0\wedge\overline{f_2(x)=0}$$

由此得$\underset{x}{M}\left[\dfrac{f_1(x)}{f_2(x)}=0\right]=\underset{x}{M}[f_1(x)=0\wedge\overline{f_2(x)=0}]$

$$\begin{aligned}&=\underset{x}{M}[f_1(x)=0]\cap\underset{x}{M}[\overline{f_2(x)=0}]\\&=\underset{x}{M}[f_1(x)=0]\cap\underset{x}{\overline{M}}[f_2(x)=0]\end{aligned}$$

$$=E_1\cap\overline{E_2}$$

(4) 方程（或不等式）组理解作这些方程（或不等式）的合取。比如，方程组

$$\begin{cases} f_1(x,\ y)=0 \\ f_2(x,\ y)=0 \end{cases}$$

理解作合取

$$f_1(x,\ y)=0\wedge f_2(x,\ y)=0,$$

而因此，方程组的解集是这个合取的真值域，或者方程组解集的交：

$$\underset{(x,y)}{M}\{f_1(x,\ y)=0\wedge f_2(x,y)=0\}$$
$$=\underset{(x,y)}{M}\{f_1(x,\ y)=0\}\cap\underset{(x,y)}{M}\{f_2(x,\ y)=0\}$$

二、引入逻辑初步知识的目的

逻辑在计算机科学和技术中的广泛运用，要求人们会对推理进行深入的分析，并把推理形式化，因为机器不能依靠直觉。现代，许多职业都必须和这样的机器交往，而能进行这样交往的人，只能是这样的人，他们掌握用公式的形式表示推理的技能，否则的话，机器不明白人的意思，人——机器系统就会出现故障。

因此，现代社会要求学生有一定的逻辑修养，要求有逻辑的分析、概括和抽象的能力，认识逻辑结构和它们之间相依关系的能力，以及分类和形式化的能力。

在中学数学教学中引入逻辑初步知识，不能单纯归结为引入逻辑符号。我们已看到逻辑运算有助于数学运算，然而，引入逻辑初步知识的目的，不仅是为了简洁地表达数学问题，更

重要的是为了发展学生的逻辑思维。

从某角度来说，教数学，首先是教推理。数学在发展学生逻辑思维方面有特殊的作用，因为在学习数学时必须进行大量的各种逻辑推理。但是，如果不给学生解释所学习的教材中的逻辑，学过之后，学生仍然不能理解这些逻辑。人们进行推理，但并不知道这些规律，这就像小孩会说中国话，但是并不知道祖国语言的语法规则一样。显然，仅仅会说中国话是不够的，祖国语言应当成为一门必修科目，这是大家公认的。同理，数学本身解决不了培养学生逻辑思维的问题。

例如，“$5\leqslant7$”是不是真命题?不少学生认为这是假命题。事实上，根据析取的准确定义可以证明答案是肯定的。因为“$5\leqslant7$”是析取命题$(5<7)\vee(5=7)$的省略写法，因为复合命题之中的一个(“$5<7$”)是真的，所以，整个析取命题是真的。

不理解析取关系，经常导致严重的运算错误。比较典型的例子是，求解诸如

$$\frac{x-2}{x-6}>0$$

之类的不等式问题。

学生都知道，这类不等式可以“分解”为两个不等式组。但是，由于忽视不等式组之间的逻辑关系，由于在“或者”的意义上误用“并且”这个连接词，学生经常得出不正确的结论，认为所给的不等式没有解；他们去找两个不等式组的解的交集，而不是用它们的解的并集作为原不等式的解集。

如果会以明显的形式运用逻辑运算，就可以用准确的语言完整地表达出解该不等式时所进行的推理，并且用比较简单的(线性的)不等式把原不等式表示为逻辑函数的形式：“当且

仅当变量x的取值能够使分子和分母都为正，或者分子和分母都为负时，分式$\frac{x-2}{x-6}$是正的"：

$$\frac{x-2}{x-6}>0 \Longleftrightarrow [(x-2>0) \wedge (x-6)>0] \vee [(x-2)<0) \wedge (x-6)<0]$$

已得到的式子可以简写为

$$[(x>2) \wedge (x>6)] \vee [(x<2) \wedge (x<6)]$$

$$\Longleftrightarrow (x>6) \vee (x<2)$$

如果学生学过逻辑初步知识，就能正确地确定原不等式的解集是能使已得到的析取命题为真的x值的集合：

$$\underset{x}{M}\left[\frac{x-2}{x-6}>0\right]=\underset{x}{M}[(x>6) \vee (x<2)]$$

$$=\underset{x}{M}(x>6) \cup \underset{x}{M}(x<2)$$

另外，当命题具有比较复杂的逻辑结构时，学生往往不会正确地作出命题的否定。

例如，下面给出加法交换律

"对于所有的a和b，$a+b=b+a$"

的一个错误证法：

"假设这个定律不成立，那么对于所有的a，b，有$a+b \neq b+a$。用$b=a$代入，就可得到$a+a \neq a+a$，就是矛盾的。因此，我们的假设是错误的，这就证明了交换律。"

对于上述证明，很多学生不能找出其中的错误，有些学生则说"错误源于用$b=a$代入"。如果学生学过逻辑初步知识，就能看出：由命题

$$(\forall a)(\forall b)(a+b=b+a) \tag{1}$$

是假的，不能推出命题

$$(\forall a)(\forall b)(a+b \neq b+a) \qquad (2)$$

是真的，(2)不是(1)的否定。命题(1)的否定是：

$$(\exists a)(\exists b)(a+b \neq b+a)$$

很多事例表明，单纯地数学学习，不能使中学生的逻辑思维得到必要的发展；必须引导中学生掌握作为人们推理的基础的逻辑规律。

斯托尼亚尔指出，在数学教学中引入逻辑初步知识的问题不求在某种特定阶段专门地孤立学习逻辑，而是在于使必要的逻辑初步知识成为数学教学本身不可缺少的一部分，必须把它们分散安排在整个数学课程中学习，成为提高数学教学在发展学生的逻辑思维方面的效果和影响的重要辅助手段。

传统形式中逻辑学与数学有本质区别，学习逻辑实质上与学习数学相脱离，是被看作与其它科目并列的学习科目。在现代的数学教学中，必须结合学习数学来学习逻辑；引进逻辑初步知识，是数学的方法、概念和语言向"逻辑"这个新的对象的自然扩展，而这种扩展又有助于更好地掌握数学方法、概念和语言本身。我国在中学数学教材中逐步渗透集合论语言的做法，就是要把它们分散到整个数学课程中，与数学教学有机地结合起来。

三、引入逻辑初步知识的可行性

国际数学教育委员会组织的第一届数学教育会（1969年）很注意考察早期发展逻辑思维的方法和形成儿童一定逻辑技能的方法。美国伊利诺大学曾编了一个用游戏加速低年级学生逻辑学习的方案，其中作者们主要注意逻辑语言的句法方面。曾在英国、澳大利亚及加拿大工作过的数学家和心理学家迪恩斯

还编了一种很有意思的学习用具《逻辑块》，利用它可以有效地用一套游戏和问题来教5—10岁儿童学习逻辑。在其中，他应用了具体东西的集合来作为早期逻辑入门的原始资料。因为儿童的智力活动是由具体集合上的运算发展到表示这些集合特征性质的命题的演算。我们小学课本里讲加法和乘法时，总是列出一些实物来引出法则，其实质就是构造具体集合。

借助集合的运算不仅可以得到数的算术运算，而且还可以得到集合特征性质的逻辑运算。

例如，集合的“并”既是抽象出数的加法的基础，也是抽象出命题的析取的基础：

$$A\cup B=C\begin{array}{l}\nearrow m(A)+m(B)=m(C)\\ \searrow A(X)\vee B(X)=C(X)\end{array}$$

$m(A)$—集合A的元素个数；$A(X)$—$X\in A$。

柯尔莫哥洛夫在1969年就指出：“没有任何根据害怕把数理逻辑的符号表示和公式广泛地引进中学……”，在以他为首编写的几何教科书中，广泛使用了集合论的思想和语言，它不仅有助于合理地建立几何理论，而且有助于克服它和中学数学的其它课程的割裂。苏联在学校数学语言现代化方面进行了积极的探索，颇见成效。在斯托尼亚尔的《数学教育学》中，可以看到他们在中、小学引入逻辑符号的一种方案。

按这个方案，规定在中、小学中逐步引用和使用下列逻辑符号：

（1）一～三年级：

$\in$——属于；

$\cup$——并；

$\cap$——交；

（按苏联教学大纲，符号“$\in$”在四年级引进，而符号“$\cup$”和“$\cap$”在五年级引进。但构思该方案的作者认为，在不远的将来，将在一～三年级学习最简单的集合论的概念和相应的语言）。

T——真命题（真）；

F——假命题（假）。

（2）四～六年级：

$\varnothing$——空集；

$\subseteq$——包含

$\overline{A}$——集合A的补；

$A\times B$——集合A乘以集合B的直积；

$A^2=A\times A$——A中元素的一切可能序偶的集合。

（3）七～八年级：

$\underset{x}{M}(p(x))$或$\{x|p(x)\}$——具有性质p的一切x的集合；

$\overline{X}$——命题X的否定；

$\wedge$——合取；

$\vee$——析取；

$\Rightarrow$——推出；

$\Leftrightarrow$——等价；

$A\xrightarrow{f}B$——集合A到集合B的映射f；

$\forall$——全称量词；

$\exists$——存在量词。

从这个方案中，可以看出他们学校数学语言现代化的大致

进程。

关于在学校数学教学中使用现代语言问题，目前争论仍很激烈，各国正在进一步的实验、研究。我们理该密切关注。

主要参考文献

〔1〕〔法〕皮埃尔·吉罗著：《符号学概论》中译本　四川人民出版社

〔2〕〔日〕池上嘉彦著：《符号学入门》　中译本　国际文化出版公司

〔3〕钱学森主编：《关于思维科学》　上海人民出版社

〔4〕徐利治、王前著《数学与思维》　湖南教育出版社

〔5〕徐利治著：《数学方法论选讲》　华中工学院出版社

〔6〕陈衡编著：《科学研究的方法论》　科学出版社

〔7〕朱智贤、林崇德著：《思维发展心理学》　北京师范大学出版社

〔8〕〔法〕雅克·阿达玛著：《数学领域中的发明心理学　中译本　江苏教育出版社

〔9〕周昌忠编译：《创造心理学》　中国青年出版社

〔10〕〔日〕米山国藏著：《数学的精神、思想和方法》　中译本　四川教育出版社

〔11〕张景中著：《数学与哲学》　湖南教育出版社

〔12〕张楚廷著：《数学与创造》　湖南教育出版社

〔13〕任樟辉著：《数学思维论》　广西教育出版社

〔14〕孙小礼著：《数学·科学·哲学》　光明日报出版社

〔15〕 莫绍揆著：《数理逻辑初步》 上海人民出版社
〔16〕〔美〕M·克莱因著：《古今数学思想》 中译本 上海科学技术出版社
〔17〕〔美〕R·柯朗、H·罗宾著：《数学是什么？》 中译本 科学出版社
〔18〕〔美〕G·波利亚著：《数学的发现》 中译本 内蒙古人民出版社
〔19〕〔美〕G·波利亚著：《怎样解题》 中译本 科学出版社
〔20〕〔苏〕A·斯托利亚尔著：《数学教育学》 中译本 人民教育出版社
〔21〕 曹才翰、蔡金法著：《数学教育学概论》 江苏教育出版社
〔22〕 郑君文、张恩华、涂荣豹编著：《数学教育学》 河海大学出版社
〔23〕 钟善基、丁尔升、曹才翰编著：《中学数学教材教法》 北京师范大学出版社
〔24〕 郑君文、张恩华著：《数学学习论》 广西教育出版社
〔25〕〔苏〕H·Я·维林金等著：《中小学数学的现代基础》 中译本 上海科技出版社
〔26〕 刘云章、马复编著：《中学数学的现代思想》 人民教育出版社
〔27〕 清华大学编：《科学技术史讲义》 清华大学出版社
〔28〕 袁小明著：《数学思想史导论》 广西教育出版社
〔29〕 梁宗巨编著：《世界数学史简编》 辽宁人民出版社

〔30〕 张奠宙、赵斌编著：《二十世纪数学史话》 知识出版社

〔31〕〔美〕H·伊乌斯著：《数学史菁华》 中译本 四川教育出版社

〔32〕〔美〕C·H·爱德华著：《微积分发展史》 中译本 北京出版社

〔33〕 王鸿钧、孙宏安著：《中国古代数学思想方法》 江苏教育出版社

〔34〕 胡世华、陆钟万著：《数理逻辑基础》 科学出版社

〔35〕 朱梧槚编著：《几何基础与数学基础》 辽宁教育出版社

〔36〕 刘云章、马复著：《数学直觉与发现》 安徽教育出版社

〔37〕 恩格斯著：《自然辩证法》 人民出版社

〔38〕 马克思著：《数学手稿》 人民出版社

〔39〕 刘凤璞、解恩泽、周民强编著：《数学中若干辩证内容简析》 人民教育出版社

〔40〕〔德〕康斯坦西·瑞德著：《希尔伯特》 中译本 上海科学技术出版社